REGULATION OF THE GENE ENCODING LIPOAMIDE DEHYDROGENASE IN *Saccharomyces cerevisiae*

BY

JOSEPH ROSS

DOCTOR OF PHILOSOPHY,

UNIVERSITY OF EDINBURGH

1989

I declare that this thesis was written and composed by myself, and that the data contained within are my own, unless otherwise stated.

Joseph Ross

April 1989

Acknowledgments

I would like first and foremost to thank my Supervisor, Ian Dawes, for all the advice and direction he has given me and for the very significant contribution he has made to the work presented here. I would also like to thank Graeme Reid, Doug Roy, Dick Dickinson, Gordon Lithgow, Al Brown, Gordon Lindsay and Lynn Kennedy, for many helpful discussions, Professor J. F. Wilkinson in whose department the work was carried out and Ian Sutherland for support in attending the 14th International Yeast Meeting. I acknowledge the receipt of a studentship grant from the Science and Engineering Research Council. In addition, I am indebted to all the undergraduate students, particularly Wendy Armstrong and Kenny McDowell, and the members of the East postgraduate laboratory for the contribution they have made both towards this work and to making my first three years in Edinburgh very happy. Finally to Julie for being there at the beginning and the end.

ABSTRACT

Lipoamide dehydrogenase is a component of the multienzyme complexes pyruvate dehydrogenase and 2-oxoglutarate dehydrogenase in *Saccharomyces cerevisiae*. A mutant, *lpd1*, lacking pyruvate dehydrogenase, 2-oxoglutarate dehydrogenase and lipoamide dehydrogenase activity had previously been isolated in yeast. In addition, this mutant had been used to screen a YEp13 based yeast gene bank leading to the isolation of a 5.5 kb region of yeast DNA containing a gene, designated *LPD*, which complements the mutation.

In this thesis the DNA sequence of 2.7 kb of the 5.5 kb region is presented. The sequenced region contains a 1.5 kb open reading frame representing the *LPD* gene. From homology between the deduced amino acid sequence of this open reading frame and the primary sequence of *E. coli* and pig heart lipoamide dehydrogenase, the *LPD* gene has been shown to represent the structural gene for lipoamide dehydrogenase. The primary sequence encoded by the *LPD* gene also shows very strong homology to several other related amino acid sequences including glutathione reductase, mercuric reductase and the lipoamide dehydrogenase sequence from *A. vinlandii* and human liver cells.

Analysis of the upstream region of the *LPD* gene led to the identification of several sites of homology between sequences within this region and known yeast regulatory motifs. These included three sites for the binding of the GCN4 protein, a sequence very similar to the UAS2 of *CYC1*, and three regions similar to the sequence TCACGTGA identified as an important element within the promoter of the *TRP1* and *GAL2* genes in *S. cerevisiae*, in the adenovirus major late promoter and as the binding site of the centromere-binding protein, CP1.

Transcript analysis of *LPD* expression during conditions of amino acid starvation was carried out. In addition, gel retardation and DNaseI protection experiments, to investigate interactions between *in vitro* synthesised GCN4 and the upstream region of the *LPD* gene, were performed. The results from these experiments suggest but do not conclusively prove that the *LPD* gene is subject to general amino acid control.

In an attempt to identify other DNA-binding proteins which interact specifically with the *LPD* gene, protein fractions from an heparin-Sepharose column were assayed by gel retardation for binding to DNA fragments from the 5' end of the gene. Several DNA-binding activities were identified including a DNA-binding protein which binds regions upstream and downstream of the translation start site. This DNA-binding activity could be competed by the addition of increasing amounts of a ligated oligonucleotide containing two copies of the TCACGTGA sequence.

The specificity and the regulatory role of the proteins identified as binding to the 5' end of the *LPD* gene remain to be determined. In this thesis a preliminary analysis of the elements involved the regulation of this gene has been carried out and future directions, in the study of transcriptional control of the *LPD* gene, discussed.

To Mum and Dad

LIST OF CONTENTS

CHAPTER 1 INTRODUCTION

The regulation of gene expression is one of the fundamental mechanisms involved in the control of cellular processes. The co-ordinated regulation of many different genes is required both throughout the various stages of the cell cycle and to enable a cellular response to enviromental changes.

Enzyme activity can be regulated by proteolytic cleavage, allosteric interaction or covalent modification. The regulation of enzyme synthesis, however, is also important in modulating the rate of metabolic reactions. For example, in the presence of glucose the expression of genes which encode enzymes involved in both the TCA cycle and oxidative phosphorylation are repressed in yeast and bacterial cells (Magasanik, 1961; Gascon *et al.*, 1968; van Wijk *et al.*, 1969). The utilisation of galactose by yeast cells, in the absence of glucose, represents another example in which the regulation of gene expression controls a metabolic pathway (Oshima, 1981).

An understanding of the molecular mechanisms involved in the regulation of gene expression is, therefore, of great importance and is currently the source of extensive investigation in both prokaryotic and eukaryotic organisms. The subject of this thesis is the investigation of the mechanisms involved in the regulation of the gene which encodes lipoamide dehydrogenase, an enzyme with at least two distinct roles in the central metabolism of the yeast *Saccharomyces cerevisiae*.

1.1 A MODEL FOR TRANSCRIPTIONAL REGULATION IN YEAST

The molecular mechanisms involved in the control of many different yeast genes have now been investigated. From this work a model has emerged which summarises the findings. This model is shown in figure 1.1. The features of this model can be divided into two distinct classes. *Cis*-acting elements which have been found, with the exception of downstream activation sites, in the 5' non-coding region of all genes, and *trans*-acting factors which interact with the regulatory elements of genes to control the rate of transcription.

1.1.1 *Cis*-acting elements

The *cis*-acting elements include the upstream activation site or UAS, operator element, TATA box, initiator element and downstream activation site or DAS. Not every gene contains all the above *cis*-acting elements but the majority, studied so far, contain at least a UAS, TATA box and initiator element (Struhl, 1987).

Upstream activation sites are required for transcription and usually determine the regulatory properties of a promoter (Guarente, 1984). They resemble mammalian enhancer sequences functioning in both orientations and at variable distances with respect to the other promoter elements and the mRNA initiation site (Chandler, 1983). Genes which are co-ordinately regulated share common UAS elements. There is now good evidence, from the study of the promoters of many different yeast genes, that UAS's are the binding sites for transcriptional regulatory proteins. A second class of UAS's exist upstream of several constitutively expressed genes. These consist of poly(dA-dT) homopolymer sequences which may function by excluding nucleosomes (Struhl, 1985).

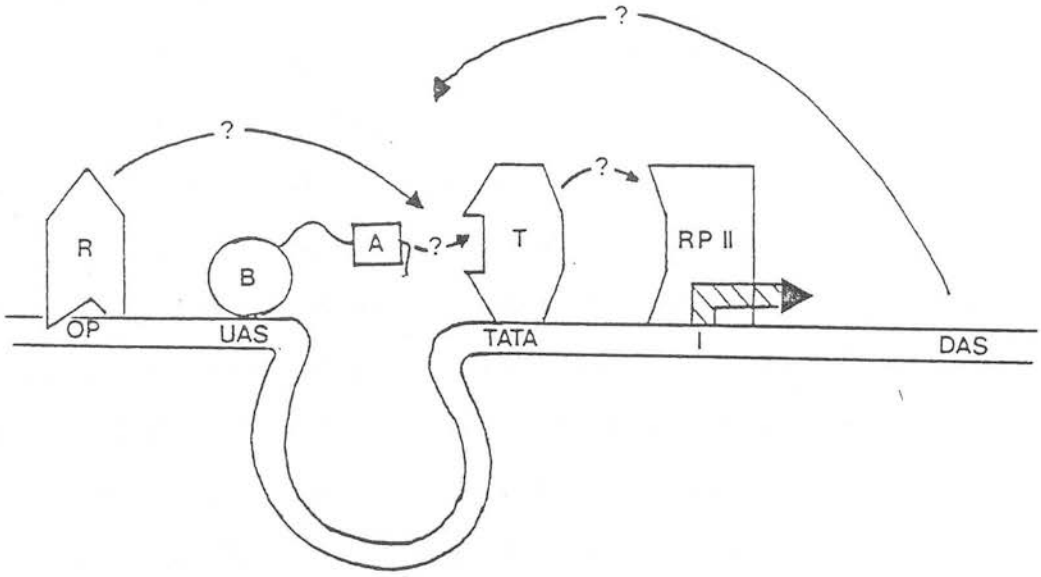


Figure 1. 1. Elements involved in transcriptional activation.

The schematic model shows elements known to have a role in transcriptional activation in yeast. OP, operator; UAS, upstream activation site; TATA, TATA element; I, RNA initiation site; DAS, downstream activation site; R, repressor; B, DNA-binding site of *trans*-acting factor; A, transcriptional activation domain of *trans*-acting factor; T, TATA-binding factor; RPII, RNA polymerase II. The hatched arrow represents mRNA transcription. The double black line represents DNA shown here looping out to allow the *trans*-acting factors to interact. The details of exactly how all these regulatory elements interact remain unclear. The arrows with question marks represent suggested interactions. Diagram modified from Struhl, (1987).

Some yeast promoters contain elements involved in the repression of transcription refered to as operator elements (Brent, 1985). An example of an operator element is found in the inducible *DAL7* gene (Cooper, 1988). The regulatory region of this gene contains three distinct classes of regulatory element. The upstream element has been resolved into two separate sites termed the upstream activation sequence, or UAS, and the upstream induction sequence or UIS . These two sites interact with an operator element termed the upstream repressor site or URS. The UIS modulates the activity of the URS which in turn, in some way, acts to inhibit the activity of the UAS. A minimum of two copies of each regulatory element are needed for the normal functioning of the *DAL7* regulatory element. How applicable this system is to the regulatory apparatus of other genes remains to be seen.

The TATA box, consensus TATAAA, is necessary but not sufficient for transcription initiation in most yeast genes, however, the absence of the TATA element from the *PGK* gene has been shown to affect the transcript start site position but not the rate of transcription (Ogden *et al.*, 1986). The *HIS3* gene has been shown to contains two functionally distinct TATA elements designated T_c and T_r involved in constitutive and inducible expression respectively (Struhl, 1986). The position of the TATA box can vary in yeast from 40 to 120 bp from the mRNA initiation site (Struhl, 1987). This is in contrast to the situation in mammalian cells in which TATA sequences are almost always located 25-30 bp from the initiation site (Breathnach and Chambon, 1981). The presence of the TATA box in many promoters indicates its function may be to bind a general transcription factor.

The initiator element controls the location of transcription initiation and may represent the binding site of either the RNA polymerase II or a protein associated with this enzyme (Burton *et al.* 1986). Removal of the TATA box and *cis*-acting elements of the *HIS4* gene and selection for supressor mutations has given rise to mutations in RNA polymerase II subunits supporting the conclusion that the *trans*-acting factors which bind to these *cis*-acting elements interact directly with RNA polymerase II (Fink, 1988). Downstream activation sites required for maximal expression, have been identified within the coding regions of the *PGK* and *PYK* genes of yeast (Mellor *et al.*, 1987; Purvis *et al.*, 1987). How these sites affect the rate of transcription remains unknown.

1.1.2 *Trans*-acting factors

Trans-acting factors which are involved in transcriptional activation and which bind to upstream activation sites have been identified in yeast. These incude GAL4, GCN4 and HAP1 proteins (Giniger *et al.*, 1985; Hope and Struhl, 1985; Pfeifer *et al.*, 1987). Analysis of truncated versions of the GAL4 and GCN4 proteins has revealed the presence of distinct DNA-binding and transcriptional activating domains (Keegan *et al.*, 1986; Hope and Struhl, 1986) in the intact proteins. The DNA-binding domains of the two proteins differ significantly. GAL4 appears to interact with DNA using the 'zinc finger' motif while the method of DNA binding for GCN4 of a 60 amino acid C-terminal region, representing the DNA-binding domain, remains unclear as it fails to conform with either the zinc finger or helix-turn-helix structure, shown to be the two major motifs involved in DNA recognition (Miller *et al.*, 1985; Pabo and Sauer, 1984).

Both proteins contain short acidic regions which display little homology. Sequential deletion analysis of a LexA-GCN4 protein, containing the LexA binding domain and the GCN4 transcriptional activator has revealed that a 19 amino acid region within the acidic region is sufficient for transcriptional activation (Hope and Struhl, 1986). Deletion analysis of the GAL4 activator has yeilded similar results (Ma and Ptashne, 1987), transcriptional activation, in this case, is localized to two short separate regions (50 - 100 amino acids).

The acidic regions show some similarity to signal sequences involved in targeting proteins to the mitochondria (von Heijne, 1986) since both types of sequence have been shown to be capable of folding into α helices. In targeting sequences this α helix is amphiphilic containing one polar and one non-polar face while for transcriptional activator sequences the putative α helix is amphipathic containing one hydrophilic face bearing acidic residues and one hydrophobic face. In addition, the GAL4 transcriptional activation domain has been replaced by a synthetic peptide which folds to give an amphipathic α helical structure (Giniger and Ptashne, 1987). The design of this synthetic peptide was based on the above observation that, although the activation regions so far examined have no obvious sequence homology, they contain amino acid sequences which, should they fold into α helices, would form negatively charged, amphipathic α helices. When expressed in yeast, the artificial protein bearing this peptide efficiently activated the *GAL1* gene which is normally activated by *GAL4*. An otherwise identical protein with the novel amino acids in a random order, which is unable to form an amphipathic structure, does not activate

GAL1 transcription. These data support the idea that the amphipathic nature of these putative helices is necessary for activity.

The size of these transcriptional activators excludes the possibility that they encode catalytic activities. Instead they are likely to represent surfaces which interact with other proteins of the transcriptional machinery such as TATA binding proteins, recently identified in yeast (Buratowski *et al.*, 1988), RNA polymerase II or other regulatory proteins. GAL4 has recently been shown to stimulate transcription in mammalian cells when GAL4 binding sites are introduced upstream of a mammalian gene (Kakidani *et al.*, 1988; Webster *et al.*, 1988). Similarly GAL4 activates transcription in both *Drosophila* and plant cells (Fischer *et al.*, 1988; Ma *et al.*, 1988). In the converse experiment the transcription of a yeast gene, *HIS3*, has been shown to be activated by the vertebrate transcription factor JUN (Struhl, 1988), which recognises a similar motif. These results indicate that a common mechanism for transcriptional activation exists in eukaryotes.

Any mechanism attempting to explain the interaction of *trans*-acting factors in the regulation of gene expression must take into consideration the distances (up to 1.4 kb) which exist between the binding sites of these regulatory proteins (Guarente, 1984). A variety of different models have been proposed whose exponents have been divided into four groups, sliders, twisters, oozers and loopers, by Ptashne (1986). The slider model proposes that regulatory proteins bind to DNA and slide along the helix, in a manner analogous to certain restriction endonucleases, to interact with other proteins. The twister model suggests *trans*-acting factors

communicate via the unwinding of intervening DNA regions. The oozer model envisages the distance between the regulatory proteins is breached by several intermediate proteins which relay the effect of one DNA-binding protein to another. Finally the looper model proposes the region of DNA between control elements loops out allowing the regulatory proteins to interact directly.

Several experiments involving the investigation of DNA-binding regulatory proteins from many different organisms indicate the looping model probably represents the way *trans*-acting factors interact. The results of key experiments which have been used to argue against the first three models discussed above and to support the looping model are summarised below.

Evidence against both the sliding and twisting models comes from the observation that the ends of the bacteriophage Mu efficiently recombine *in vitro* under the direction of site-specific recombinational proteins when the two binding sites are present on totally separate intertwined DNA circles (Craigie and Mizuuchi, 1986). The proteins involved can therefore not interact in this case through unwinding of the DNA or by sliding along the DNA as the two *cis*-acting elements are on totally separate DNA molecules. Plon and Wang, (1986), designed an experiment that topologically separates the DNA site recognised by a regulatory protein from a gene without destroying gene activation. They constructed a 'tailed circle' in which the regulatory sequence, the SV40 enhancer, forms a hairpin protruding from an otherwise intact circular DNA molecule that includes a gene (human β-globin). When this construct is introduced into cells, transcription of the gene is enhancer dependent, as it is when gene and enhancer are present on ordinary

DNA molecules. Thus, twisting of the enhancer cannot be responsible for gene activation, because twisting the protruding enhancer could have no effect on the topology of the gene. The oozing model seems implausible given the large distances often present between *cis*-acting elements and experiments performed *in vitro* show that for the case involving the ends of Tn*3*, the amount of protein sufficient to catalyse the reaction is insufficient to cover the DNA between the sites (Krasnow *et al.*, 1983).

The evidence for looping includes the following observations. First, although *cis*-acting elements can fuction at variable distinces from each other, in several cases it has been demonstrated that the ability of these elements to interact is dependent on the spacing between them remaining an integral number of helical turns (Dunn *et al.*, 1984; Martin *et al.*, 1986). If the spacing is varied by half-integral numbers of turns the interaction of the elements is reduced suggesting the DNA can bend but not twist to allow interaction of proteins binding to these *cis*-acting elements. Second, the DNase sensitivity pattern of DNA between two *cis*-acting elements alters when proteins are bound to the elements. The DNA between the sites becomes alternately hypersensitive and resistant to DNase and the interval separating these recurring hypersensitive and resistant sites is five base pairs (Horschild and Ptashne, 1986). This pattern of DNase sensitivity can be explained by looped DNA which will be more sensitive to DNase where the minor groove, the site recognised by DNase, is expanded and more resistant when the minor groove is compressed. Finally, electron microscopy has now been used to provide direct visual evidence for DNA looping induced by λ repressor binding (Griffith *et al.*, 1986).

The above results together strongly support the proposed looper model for transcriptional activation, however, further investigation of the characteristics of yeast promoters is needed before the looping model can be confidently proposed as the correct mechanism in *S. cerevisiae*

Thus, in the model shown in figure 1.1, a specific *trans*-acting factor binds to the UAS. It then mediates its effect by interacting with other regulatory proteins, including the TATA binding protein, and ultimately RNA polymerase II. This interaction is made possible by the looping out of the intervening DNA. Thus the DNA acts as a scaffold for the assembly of an active transcription complex.

The regulatory regions of yeast genes, encoding proteins involved in galactose utilisation, nitrogen assimilation and degradation, oxidative phosphorylation, mating type control and glycolysis have all been investigated by other workers. A theme emerging from this work is the high level of complexity, often present, with several upstream *cis*- and *trans*-acting elements involved in controlling the transcription of each gene. As yet, however, very little work has been done to investigate the regulation of the genes which encode enzymes involved in the TCA cycle. This is despite the central metabolic position of this cycle which has key roles in both energy synthesis and the production of precursors for a wide variety of biosynthetic pathways. An investigation of the regulation of the gene encoding lipoamide dehydrogenase provides an opportunity to examine the control mechanism of a gene encoding a protein involved in two metabolic steps which are important in the functioning of the TCA cycle.

1.2 LIPOAMIDE DEHYDROGENASE ENZYMOLOGY AND REGULATION

Lipoamide dehydrogenase is a component of the two multienzyme complexes pyruvate dehydrogenase and 2-oxoglutarate dehydrogenase which catalyse the oxidative decarboxylation of pyruvate to acetyl CoA and 2-oxoglutarate to succinyl CoA respectively (Reed, 1974). The metabolic position of these two multienzyme complexes is shown in figure 1.2. Both enzyme complexes are required, during growth on non-fermentable carbon sources, for the synthesis of ATP. The role of the two complexes in the TCA cycle dictates that they are also important in producing substrates for anabolic pathways such as glutamate and porphyrin biosynthesis. All available evidence suggests both multienzyme complexes function within the mitochondrial matrix-inner membrane space.

The two complexes mediate an analogous series of reactions catalysed by multiple copies of three types of enzyme activity: pyruvate or 2-oxoglutarate dehydrogenase (E1); acetyl or succinyl transferase (E2) and lipoamide dehydrogenase (E3). The E1 and E2 components are specific to their respective complexes while lipoamide dehydrogenase has been shown to be functionally interchangable between both (Mukherjee, 1965). In *Escherichia coli* and *Saccharomyces cerevisiae* this is encoded by a single nuclear gene (Guest and Creaghan, 1973; Dickinson *et al.*, 1986).

The sequence of reactions catalysed by the two multienzyme complexes is summarised in figure 1.3. The dehydrogenase component, E1, mediates the thiamine pyrophosphate-dependent decarboxylation of the 2-oxoacid and its subsequent oxidation to an acyl group by simultaneous acylation of the lipoamide coenzyme of the transacylase

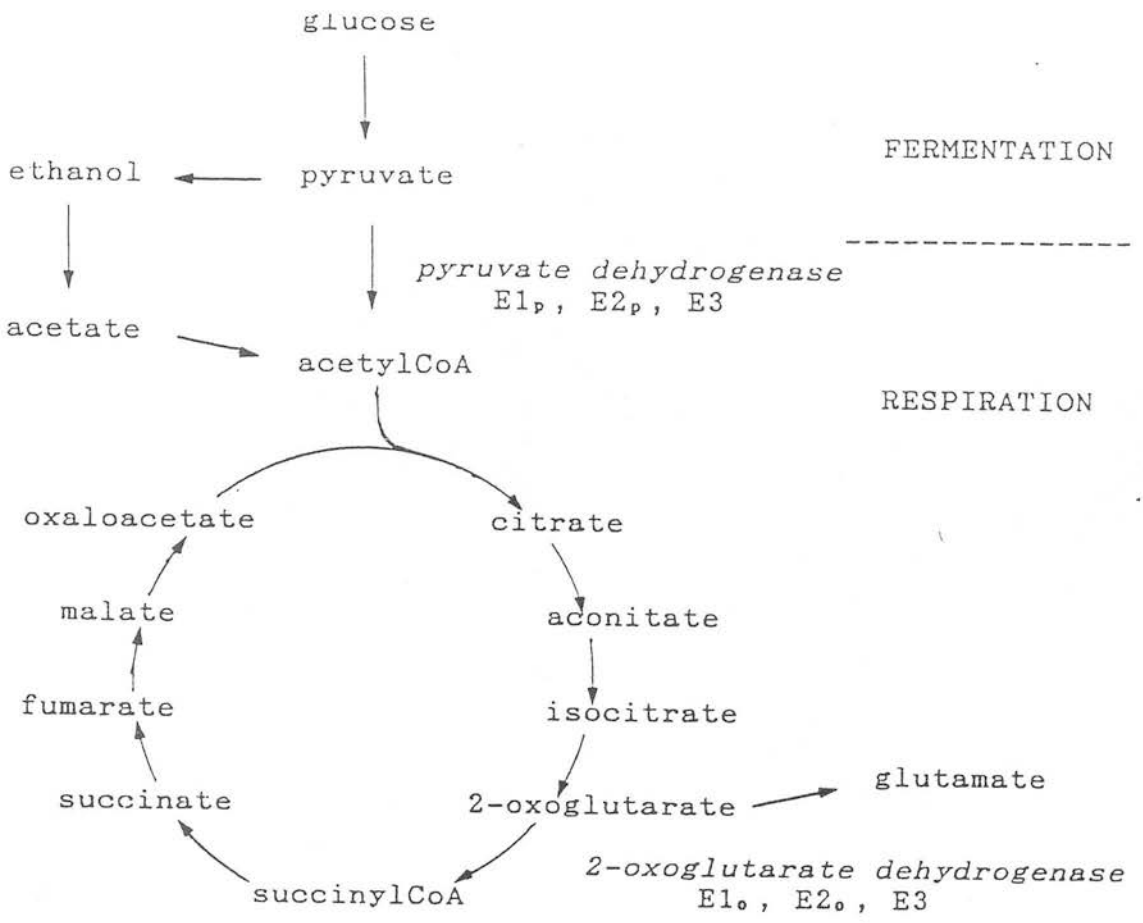
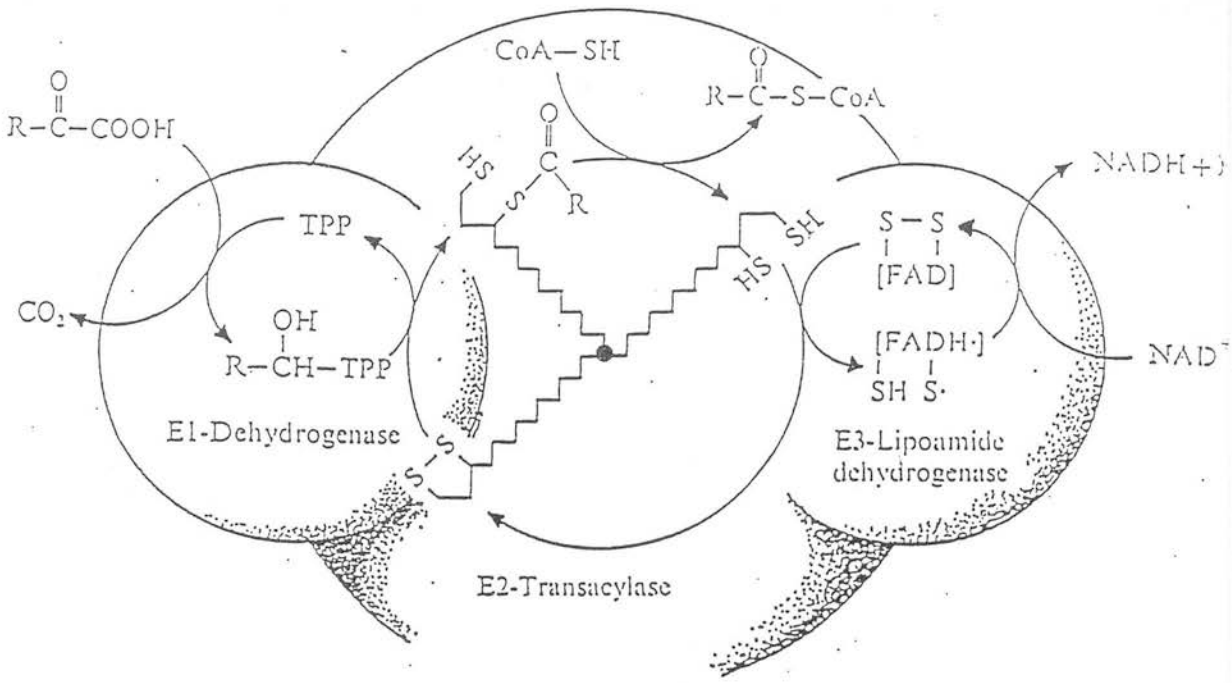


Figure 1. 2. Role of lipoamide dehydrogenase in yeast metabolism.

When grown on a fermentable substrate, such as glucose, *Saccharomyces cerevisiae* ferments the sugar to ethanol, and the TCA cycle is repressed by catabolite repression although there is still a requirement for some activity for the synthesis of metabolites such as haem and the glutamate family of amino acids (glutamate, glutamine, proline, arginine and lysine).

When grown on non-fermentable substrates, such as glycerol and lactate, the substrate is respired, entering the TCA cycle via pyruvate, and both pyruvate and 2-oxoglutarate dehydrogenase are required. On ethanol and acetate there is a reduced need for the pyruvate dehydrogenase complex. E1, E2 and E3 refer to the three enzymic components of the complexes.



$$R.CO.COOH + CoASH + NAD^{+} \rightarrow R.CO \sim SCoA + CO_2 + NADH + H^{+}$$

Figure 1. 3. Diagrammatic representation of the sequence of reactions catalysed by the E1, dehydrogenase, E2, dihydrolipoyl transacylase and E3, lipoamide dehydrogenase components of the pyruvate and 2-oxoglutarate dehydrogenase complexes.

The lipoamide coenzyme, joined to the polypeptide chain (●) of a transacylase component (E2) by amide linkage to the ε-amino group of a lysyl-residue, is shown interacting at three active sites in oxidised, acylated-reduced, and reduced states. R, CH_3- or $COOH.CH_2.CH_2$-: TPP, thiamine pyrophosphate; CoA-SH, coenzyme A; FAD, flavin adenine dinucleotide; NAD^+ and NADH, nicotinamide adenine dinucleotide, oxidised and reduced form.

component, E2. The acyllipoamide possesses an energy-rich thioester bond that is conserved when the acyl group is transferred to CoA by the transacylase component. Finally, the reduced lipoamide coenzyme is reoxidised by the flavoprotein component lipoamide dehydrogenase, E3, with the concomitant reduction of a molecule of NAD^+ (Williams *et al.*, 1976).

The organisation of the components of pyruvate dehydrogenase and 2-oxoglutarate dehydrogenase are considerably more complex than figure 1.3 might suggest. The stoichiometry and organisation of the three components has been studied in a number of different organisms (Reed, 1974) and is summarised in table 1.1 and figure 1.4. The assembly of the *E. coli* pyruvate dehydrogenase transacetylase and the transacylase components from both *E. coli* and mammalian 2-oxoglutarate dehydrogenases is based upon an octahedral 432 symmetry. This forms the structure onto which the E1 and E3 components, each in dimeric form, are positioned.

The mammalian pyruvate dehydrogenase complex has a different structure. A pentagonal dodecahedral arrangement of transacetylase components forms the central structure onto which the E1 and E3 components are located. The pyruvate dehydrogenase component has the subunit composition $\alpha_2\beta_2$. The α chain catalyses the decarboxylation of pyruvate while the β chain catalyses the reductive acetylation of the lipoyl moiety of the transacetylase. Lipoamide dehydrogenase again assembles in dimeric form. The mammalian pyruvate dehydrogenase complex also contains kinase and phosphatase subunits involved in regulating the enzyme activity of the complex (Reed, 1974). The presence of a further component, termed subunit X, has also been reported for the mammalian pyruvate

Table 1. 1. Subunit composition of *E. coli* and mammalian pyruvate and 2-oxoglutarate dehydrogenase complexes.

PDC, pyruvate dehydrogenase complex; OGDC, 2-oxoglutarate dehydrogenase complex. Diagram from Reed, (1974).

Enzyme	Subunit assembly	Subunit M_r	Subunits per complex
E. coli PDC			
Pyruvate dehydrogenase	2	96,000	24
Transacetylase	24	70,000	24
Lipoamide dehydrogenase	2	56,000	12
E. coli OGDC			
2-oxoglutarate dehydrogenase	2	95,000	12
Transsuccinylase	24	42,000	24
Lipoamide dehydrogenase	2	56,000	12
Mammalian PDC			
Pyruvate dehydrogenase (α)	2	41,000	40
Pyruvate dehydrogenase (β)	2	36,000	40
Transacetylase	60	52,000	60
Lipoamide dehydrogenase	2	55,000	10
Kinase	1	50,000	3
Phoshatase	1	100,000	5
Mammalian OGDC			
2-oxoglutarate dehydrogenase	2	95,000	12
Transsuccinylase	24	42,000	24
Lipoamide deyhdrogenase	2	56,000	12

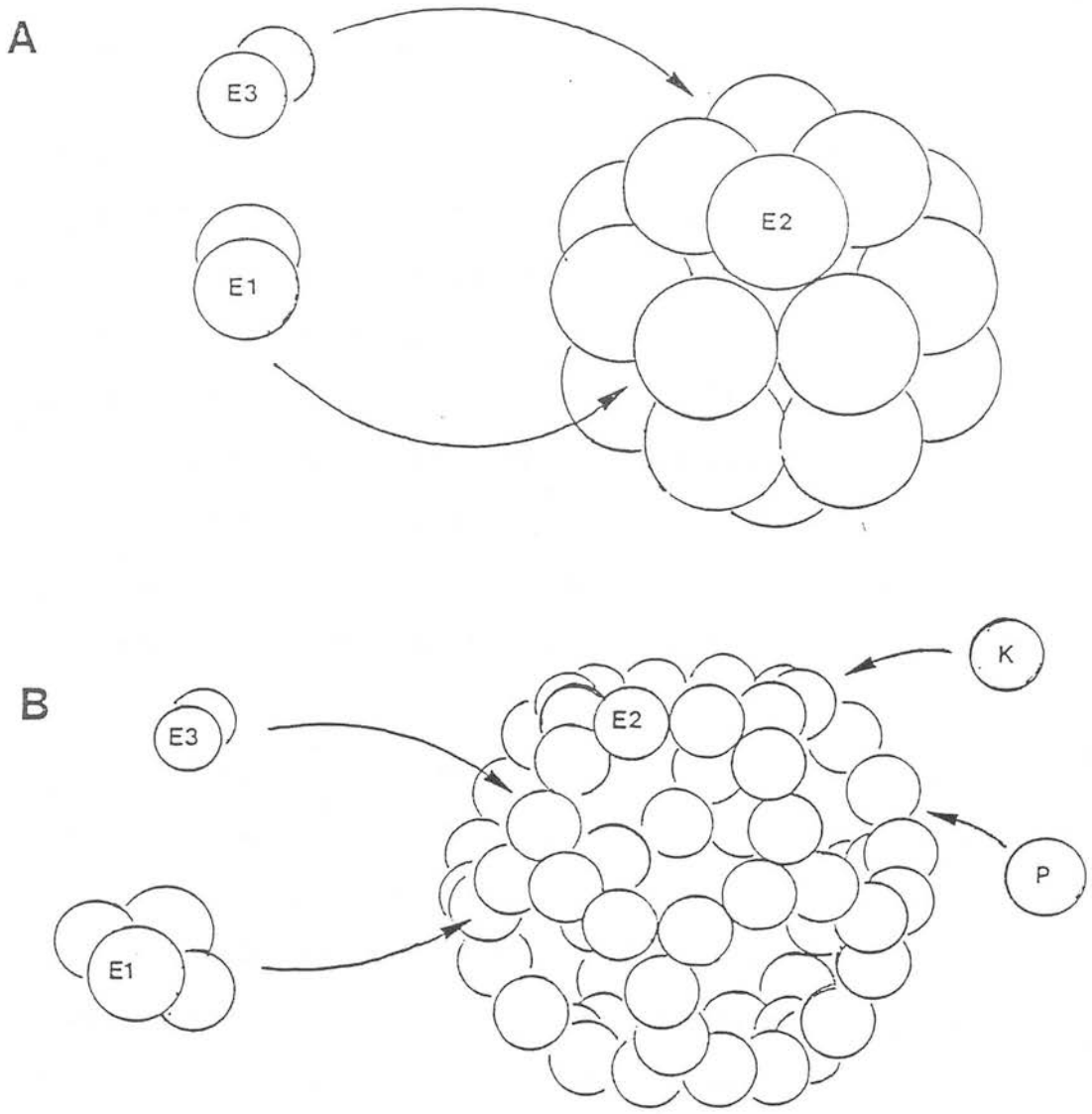


Figure 1. 4. Quaternary structure of pyruvate dehydrogenase and 2-oxoglutarate dehydrogenase complexes.

A. Structure of *E. coli* pyruvate dehydrogenase, *E. coli* 2-oxoglutarate dehydrogenase, and mammalian 2-oxoglutarate dehydrogenase complexes. The 24 E2 subunits form into a truncated cube. The 12 E1 dimers assemble on the twofold positions (*i.e.*, on the edges) of the transacylase cube, and the 6 E3 dimers assemble on the fourfold positions (*i.e.*, on the faces).

B. Structure of mammalian pyruvate dehydrogenase complex. The 60 E2 subunits form a pentagonal dodecahedral structure. The 20 E1 tetramers are located at the twofold positions (*i.e.*, the edges) and the 5 E3 dimers are located at the fivefold positions (i.e., the faces). The kinase and phosphatase subunits bind to the transacetylase subunits.

dehydrogenase complex. This component, previously believed to be a proteolytic fragment of E2, has recently been shown to represent a distinct polypeptide which contains lipoic acid and which may be involved in the transport of acetyl groups within the complex (Hodgson *et al.*, 1986).

The structural integration of the different components, which mediate the oxidative decarboxylation of pyruvate and 2-oxoglutarate, increases the overall reaction rates and minimises side reactions which might interfere with the intended catalytic step (Bates *et al.*, 1977).

The mechanisms by which the two complexes are regulated have been investigated and reflect their structural similarity (Reed, 1974). Both complexes are subject to product inhibition. Thus, acetyl CoA and NADH inhibit pyruvate dehydrogenase (Reed, 1969) while succinyl CoA and NADH inhibit 2-oxoglutarate dehydrogenase (Garland, 1964). These inhibitory effects are reversed competitively by CoA and NAD^+. The sites of acyl CoA and NADH inhibition are the transacylase and flavoprotein components respectively (Schwartz and Reed, 1970; Parker and Weitzmann, 1972).

The two complexes are also subject to feedback inhibition. The reactions mediated by the complexes result in the generation of ATP. It is, therefore, not suprising that the activity of pyruvate dehydrogenase and 2-oxoglutarate dehydrogenase are regulated by the phosphorylation state of the nucleotide pool (Atkinson, 1968). In this case the energy charge of the cell regulates the activity of the E1 dehydrogenase component of both complexes.

The pyruvate dehydrogenase complex, in eukaryotic organisms, is also subject to regulation by covalent modification. The kinase

and phosphatase subunits, mentioned earlier, have been found in a wide variety of organisms including *S. cerevisiae* (Linn *et al.*, 1969). The regulation by covalent modification is summarised in figure 1.5. The two types of regulatory subunits bind to the transacetylase component of the complex. The sites of phosphorylation are three seryl residues in the α chain of the pyruvate dehydrogenase component of the complex (Barrera, 1972). As shown in figure 1.5 the activity of the kinase and phosphatase subunits are themselves regulated by a variety of metabolic intermediates and divalent cations (Reed, 1974).

The three different control mechanisms, described above, tightly regulate the activity of pyruvate dehydrogenase and 2-oxoglutarate dehydrogenase. This reflects the important roles these two complexes play in the modulation of metabolic flux through the TCA cycle.

Roles for lipoamide dehydrogenase have been demonstrated in several other areas of metabolism. In mammalian cells the multienzyme complexes that catalyse the oxidative decarboxylation of the branched chain 2-oxoacids derived by transamination from leucine, valine and isoleucine have been shown to contain lipoamide dehydrogenase (Dancis *et al.*, 1972; Goedde and Keller, 1967) so far, however, there has been no report of the occurrence of this activity in yeast. The reversible oxidative decarboxylation of glycine has also been shown to involve lipoamide dehydrogenase in the aerobic bacterium *Arthrobacter globiformis* (Kochi and Kikuchi, 1976), in the anaerobe *Peptococcus glycinophilus* (Robinson *et al.*, 1973) and in rat liver mitochondria (Kochi and Kikuchi, 1976). A role for lipoamide dehydrogenase has also been postulated for the

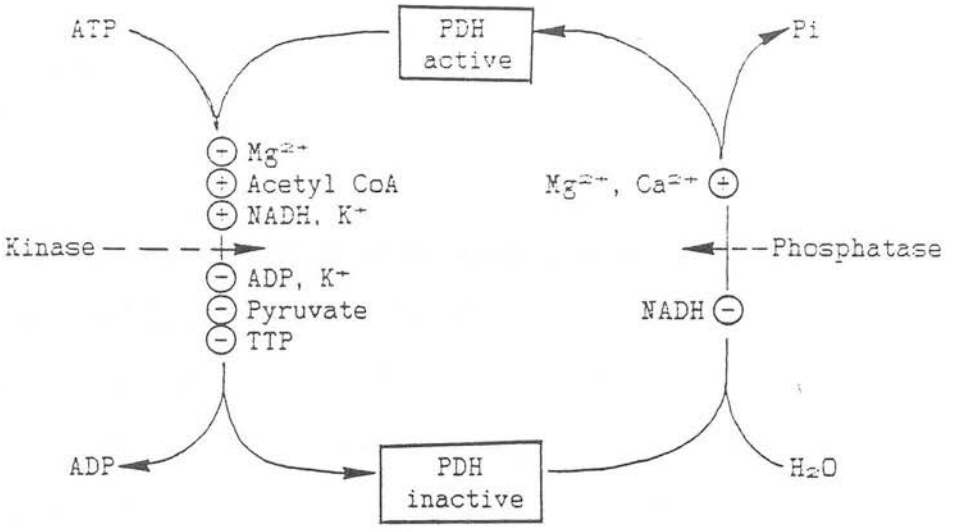


Figure 1. 5. Schematic representation of the covalent modification of pyruvate dehydrogenase and its control by various metabolites.

Diagram from Reed, (1974).

NAD^+-dependent conversion of lactate to pyruvate by *Butyribacterium rettgeri* (Wittenberger, 1964).

1.3 ORGANISATION OF THE GENETIC INFORMATION FOR THE PYRUVATE AND 2-OXOGLUTARATE DEHYDROGENASE COMPLEXES IN *E. coli*

Genetic analysis in *E. coli* has revealed that the genes encoding the subunits of pyruvate dehydrogenase and 2-oxoglutarate dehydrogenase exist as two polycistronic operons (Guest and Rice, 1984; Spencer and Guest, 1985). The *ace* operon contains the E and F genes which encode the E1 and E2 components, respectively, of the pyruvate dehydrogenase complex. A third gene, *lpd*, which is linked to the *ace* operon but which can operate under its own promoter encodes lipoamide dehydrogenase. The *suc* operon contains the A, B, C and D genes which encode the E1 and E2 components of the 2-oxoglutarate dehydrogenase complex and the β and α components of the succinyl-CoA synthetase complex respectively (Miles and Guest, 1987). Figure 1.6 summarises the transcripts of the *suc*, *ace* and *lpd* genes. S1 nuclease mapping of the *suc*, *ace* and *lpd* transcripts has revealed that most of the E3 components supplying the pyruvate dehydrogenase complex are synthesised from an *aceEF-lpd* read through transcript (Spencer and Guest, 1985). Additional *aceEF* transcripts, terminating after the *aceF* gene, provide a transcriptional basis for the observed stoichiometric excess of the E1 and E2 components, relative to E3, in the assembled pyruvate dehydrogenase complex.

Analysis of the relative enzyme activities of pyruvate dehydrogenase and 2-oxoglutarate dehydrogenase show that in *E. coli*

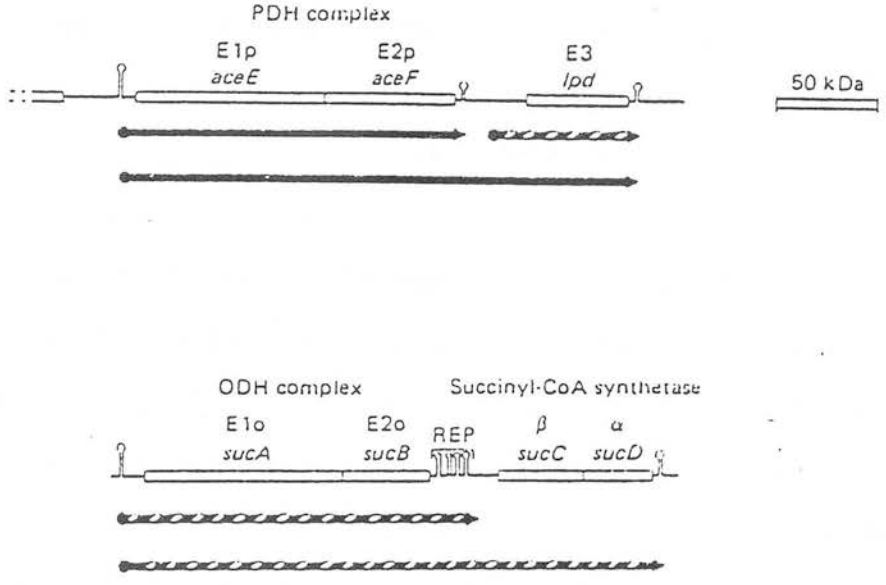


Figure 1. 6. Summary of the transcripts of the *E. coli suc*, *ace* and *lpd* genes.

The coding regions are shown as open boxes and regions of potential secondary structure associated with putative promoters, terminators and REP's (repeated extragenic palindromes) are shown as stem loops. Transcripts synthesised co-ordinately with the 2-oxoglutarate dehydrogenase complex are shown as hatched arrows and those synthesised during expression of the pyruvate dehydrogenase complex as black arrows. Diagram from Miles and Guest, (1987).

they can vary from 4:1 to 1:2 in cells grown on different carbon sources (Guest and Creaghan, 1973). The organisation of the transcripts for the two complexes provides a mechanism which allows the cell to co-ordinate synthesis of the different multienzyme components in response to growth conditions. Growth on pyruvate, which stimulates pyruvate dehydrogenase activity, results in elevated levels of *aceEF* and *aceEF-lpd* transcripts while growth on acetate, which supresses pyruvate dehydrogenase activity, results in increased levels of *sucABCD* and *lpd* transcripts. Thus, the biosynthesis of lipoamide dehydrogenase can be co-ordinated with that of the other components of the two multienzyme complexes.

The absence of polycistronic operons in eukaryotic organisms dictates that the regulation of expression of the genes encoding the pyruvate dehydrogenase and 2-oxoglutarate dehydrogenase subunits will be significantly different from the situation in *E. coli*.

1.4 WORK LEADING TO AN INVESTIGATION OF THE GENE ENCODING LIPOAMIDE DEHYDROGENASE IN *Saccharomyces cerevisiae*

The investigation of the regulation of the gene encoding lipoamide dehydrogenase *Saccharomyces cerevisiae* stems, originally, from observations made during the analysis of the early events involved in the switch from vegetative growth by mitosis to meiotic growth during the initiation of sporulation. Mutants derepressed for sporulation were apparently affected in mitochondrial function (Vezinhet *et al.*, 1979) Analysis of acetate metabolism using ^{13}C NMR spectroscopy showed that one of the early events in sporulation was a reduction in activity of the TCA cycle and an

increase in the conversion of acetate to glutamate (Dickinson *et al.*, 1983). Derepressed sporulation mutants were shown to accumulate, during vegetative growth, levels of glutamate typical of sporulating cells (Dickinson *et al.*, 1985). 2-oxoglutarate dehydrogenase and succinyl CoA synthetase represent potential control points, for this switch, as they are involved in the TCA cycle but not the glyoxylate cycle. Analysis of 2-oxoglutarate dehydrogenase activities in wild type and sporulation derepressed mutants revealed distinct differences between the two (Dickinson *et al.*, 1985).

On the basis of the above results, mutants lacking 2-oxoglutarate dehydrogenase activity were isolated using the observation that a mutant lacking this activity could grow weakly on ethanol but not on glycerol (Subik, 1972).

A mutant lacking 2-oxoglutarate dehydrogenase activity but which also lacked pyruvate dehydrogenase and lipoamide dehydrogenase activity was isolated (Dickinson *et al.*, 1986). It was, therefore, likely that the mutation represented a lesion in a gene involved in the biosynthesis of lipoamide dehydrogenase. The mutant, designated *lpd1*, may have affected the synthesis of an essential cofactor of lipoamide dehydrogenase rather than the structural gene for the enzyme. This was unlikely since haploids carrying the *lpd1* mutation did not respond to exogenous lipoamide or FMN. Succinate-cytochrome C oxidoreductase activity was also relatively unaffected in the mutant ruling out the possibility of a requirement for flavin prosthetic groups.

The *lpd1* mutant was used to screen a yeast gene bank, based on the multicopy vector YEp13, for sequences which complemented the lesion. A number of sequences were isolated which all contained a

common 3.7 kb *Xho*I restriction fragment, designated *LPD1* (later renamed *LPD*) Transformed cells carrying the plasmid pGP1, a derivative of YEp13 with a 5.5 kb yeast fragment insert, containing the *LPD* gene, produced ten times the wild type level of lipoamide dehydrogenase. From these results it was suggested that the *LPD* gene encoded either the structural gene for lipoamide dehydrogenase or a gene product that regulates the synthesis of this enzyme.

Two pieces of evidence supported the idea that *LPD* was indeed the structural gene for lipoamide dehydrogenase. Firstly, the *lpd1* mutation was recessive so if the *LPD* gene did have a regulatory function it must be acting in a positive way. Secondly, diploids heterozygous for the *lpd1* mutation had approximately half the level of lipoamide dehydrogenase, found in the wild type diploid, but the normal amount of fumarase and NAD^+-linked isocitrate dehydrogenase (Dickinson *et al.*, 1986). Such a gene dosage effect, while consistent with *LPD* encoding the structural gene for lipoamide dehydrogenase, can not easily be reconciled with the concept of a positive regulatory function.

The *LPD* sequence was integrated into the yeast genome and shown to represent the gene in which the initial *lpd* mutation had been made (Roy and Dawes, 1987). Transcript analysis of the *LPD* gene, during growth on different carbon sources, also revealed that the gene was subject to catabolite repression. This was expected if *LPD* represents the structural gene for lipoamide dehydrogenase since the enzyme was known to be subject to this form of regulation from enzyme activity measurements and from the use of E3 antibodies to monitor the levels of the enzyme during the switch from fermentable to non-fermentable carbon sources (Roy and Dawes, 1987).

An investigation of the regulatory mechanisms involved in controlling the rate of expression of the *LPD* gene is of interest for several reasons. The gene encodes an enzyme which is a component of two multienzyme complexes which play key roles in the functioning and regulation of the TCA cycle. As discussed earlier, there is still very little known about the regulation of the genes which encode enzymes involved in this cycle. The role of lipoamide dehydrogenase in two separate complexes also raises questions concerning the mechanism for the co-ordinated biosynthesis of the E3 subunit with those of the E1 and E2 subunits of the pyruvate dehydrogenase and 2-oxoglutarate dehydrogenase multienzyme complexes.

Lipoamide dehydrogenase is encoded in the nucleus, synthesised in the cytoplasm and functions in the mitochondrial matrix/inner membrane space. The mechanism of targeting of the enzyme to its correct subcellular location is therefore also of interest.

The *LPD* gene has been shown to be subject to catabolite repression (Roy and Dawes, 1987). Analysis of the regulation of this gene represents an opportunity to investigate this control network which remains poorly understood in yeast (Entian, 1986). Finally, regulation of the *LPD* gene may have a role in the regulation of 2-oxoglutarate dehydrogenase during the initiation of sporulation.

In order to determine the exact nature of *LPD* the first aim of this work was to sequence the gene. This would clearly establish whether or not *LPD* was the structural gene for lipoamide dehydrogenase. A comparison of the primary structure of *E. coli* lipoamide dehydrogenase, derived from the DNA sequence of the *lpd*

gene (Stephens *et al.*, 1983), with pig heart lipoamide dehydrogenase amino acid sequences, obtained from tryptic peptides (Williams *et al.*, 1982), revealed that a high level of homology exists between the two. This is in agreement with the reported low immunogenicity of the E3 component suggesting that the enzyme has been highly conserved throughout the course of evolution (De Marcucci, 1985). The amino acid sequence deduced from the *LPD* gene should exhibit a high degree of homology with the two established lipoamide dehydrogenase primary sequences if the gene is indeed the structural gene for lipoamide dehydrogenase in yeast.

Sequencing the 5' non-coding region of the *LPD* gene represented another important objective of this work. This region should contain most of the regulatory elements of the gene and would therefore represent the sequence of greatest interest in the examination of the elements involved in modulating the expression of the gene. A search for sites of homology between this region and sequences identified as important regulatory sites within other yeast genes would also be possible once the DNA sequence was determined. The relevance of such a search has become increasingly important as more regulatory consensus sequences have been identified and the presence of common regulatory elements in a wide range of genes has been demonstrated.

An investigation of the elements involved in the regulation of the *LPD* gene, using the range of techniques recently developed in the study of other genetic regulatory elements, represented the next logical step in this work. Gel retardation and DNA footprinting techniques are now established in the study of DNA/protein interactions involved in the regulation of gene expression (Fried

and Crothers, 1981; Garner and Revzin, 1981; Galas and Smith, 1978). Both techniques should help determine the location of regulatory elements within the *LPD* gene. The identification of specific *cis*-acting elements in the upstream region of the *LPD* gene should also provide some direction in the choice of conditions used to examine the level of *LPD* expression.

CHAPTER 2 MATERIALS AND METHODS

2.1 STRAINS AND PLASMIDS

2.1.1 Yeast Strains

The strains of *Saccharomyces cerevisiae* used are listed below:

Table 2.1: *S. cerevisiae* strains used

STRAIN	GENOTYPE	SOURCE
D273-10B ATCC 24657	*MATα* wild type	G. A. Reid
DC5	*MATa his3 leu2-3 leu2-13 can1-11*	I. W. Dawes

2.1.2 Bacterial Strains

The strains of *Escherichia coli* used are listed below:

Table 2.1: *E. coli* strains used

STRAIN	GENOTYPE	SOURCE
HB101	F^- *recA13 hsdS20 ara14 proA2 lacY1 galK2 rpsh20 xyl15 mt1-1 supE44* (Sm^r) λ^-	I. W. Dawes
MM294	*pro- endoA- thi- hsdR- hsdM*	G. A. Reid
JM101	*Δ(lac-pro) supE thi/F' traD36 proAB* $lacI^q$ *ZΔM15*	P. Fantes

2.1.3 Cloning vectors

The integration vector pGP-R1, a derivative of the YEP13 based vector pGP1 (Roy and Dawes, 1987), was used to isolate the 3.7 kb fragment for shotgun cloning into M13 sequencing vectors (Messing, 1983). pGP1 and pGP-R1 are shown in figure 2.1. The expression vector pSP64-GCN4 (Hope and Struhl, 1985), shown in figure 2.2, used for *in vitro* transcription/translation of GCN4 protein, was obtained from I. Hope and K. Struhl. A 1.4 kb *SalI/KpnI* fragment of the *LPD* gene was ligated into the polylinker of pUC18 producing the vector pID3 (unpublished results). This plasmid was used in the isolation of DNA fragments required for gel retardation and DNaseI footprinting experiments. pID3 is shown in figure 2.3.

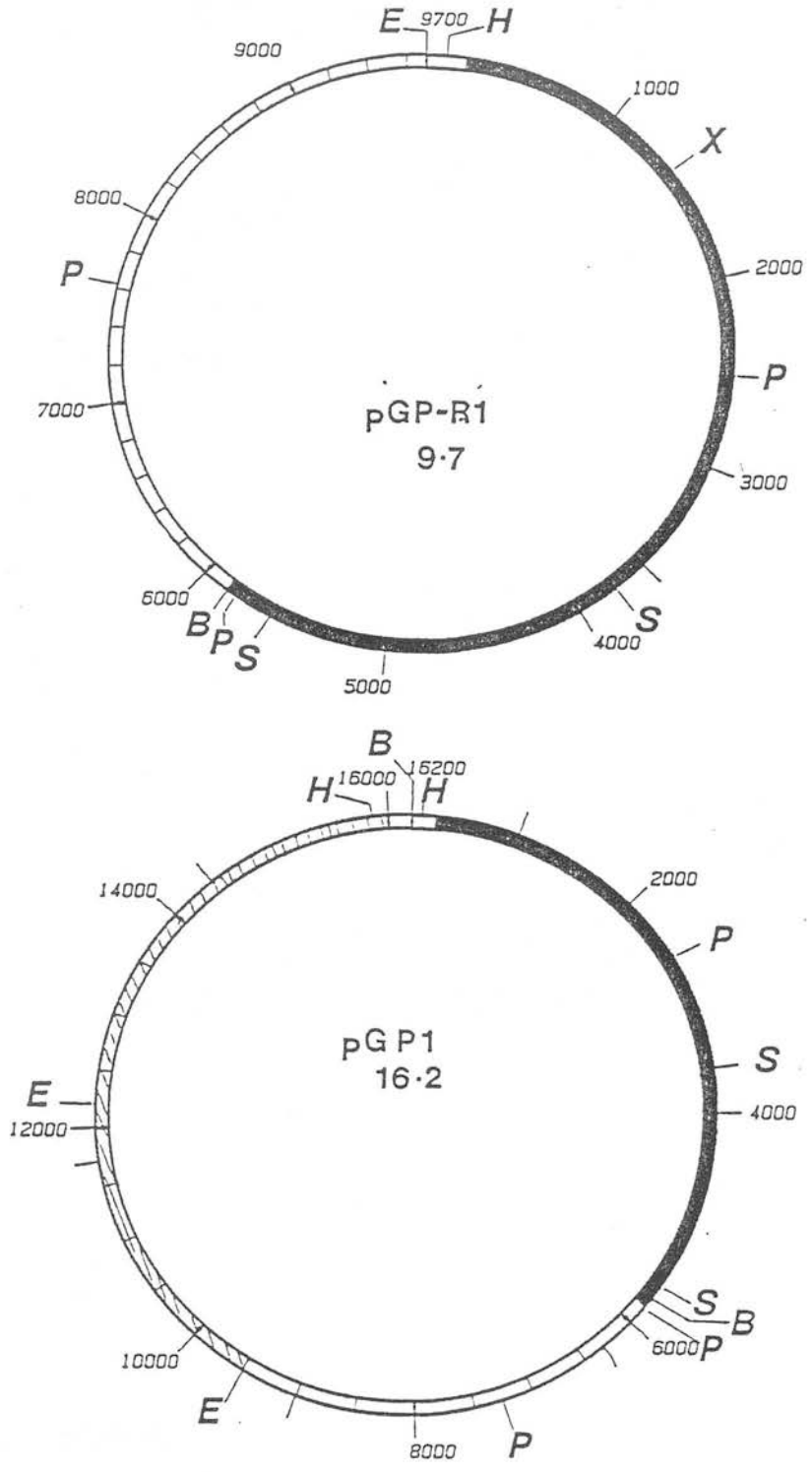


Figure 2. 1. Plasmids pGP1 and pGP-R1.

■, Insert DNA; ▨, yeast *LEU2* and 2 μm vector sequences; □, sequences from the bacterial plasmid pBR322. Restriction sites are: *Pvu*II (P); *Sst*II (S); *Eco*RI (E); *Hin*dIII (H); *Bam*HI (B); and *Xho*I (X).

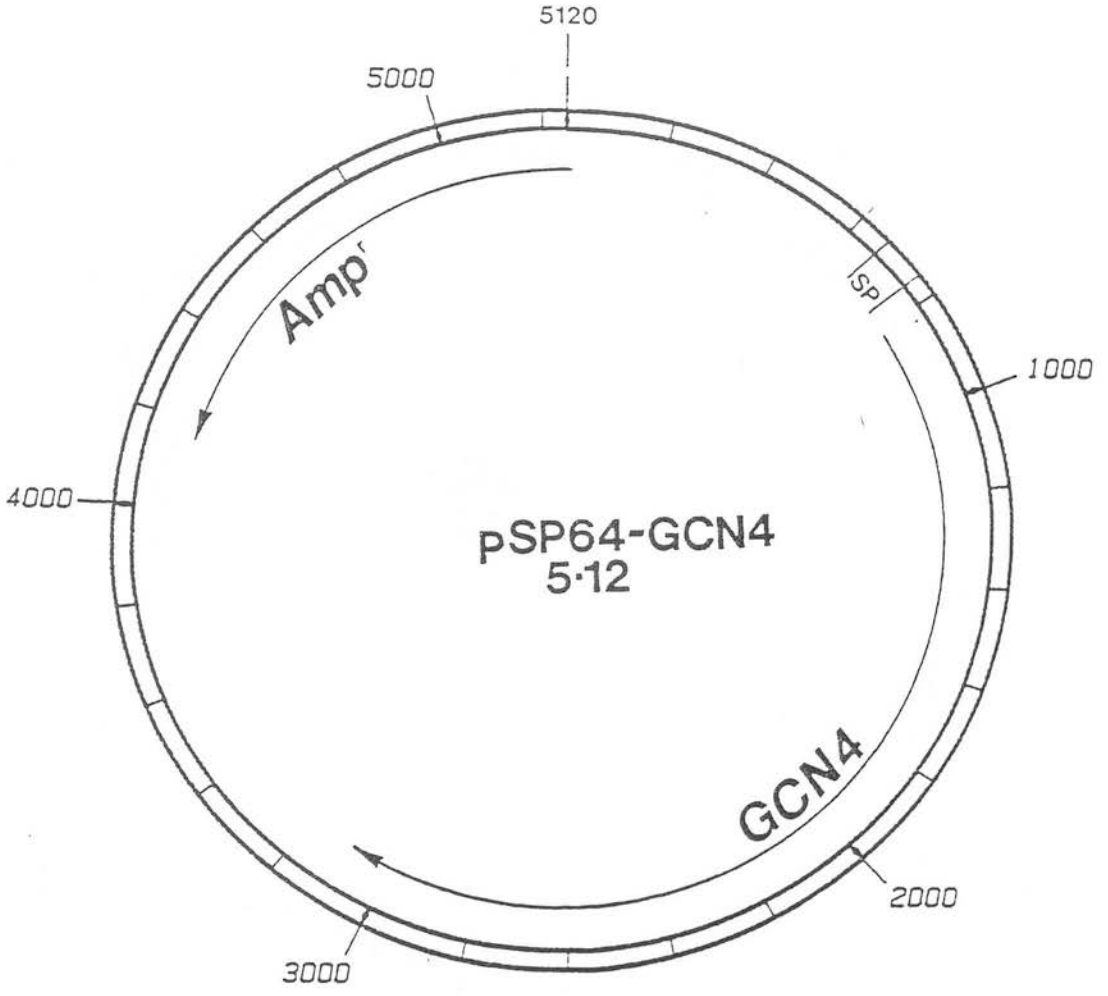


Figure 2. 2. pSP64-GCN4

Location and direction of the *GCN4* and ampicillin resistance genes is shown. SP: SP6 promoter.

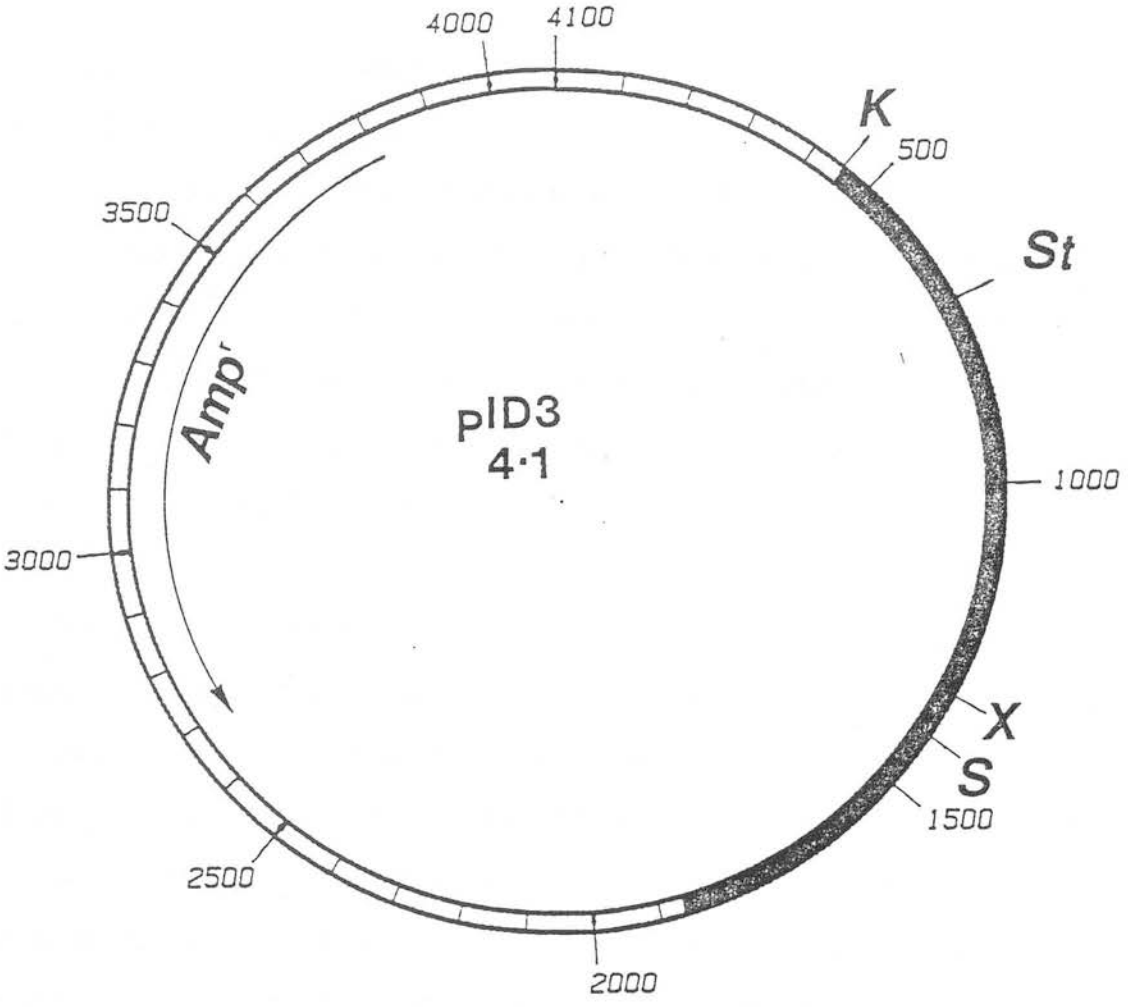


Figure 2. 3. pID3

Location and direction of the ampicillin resistance gene is shown. This plasmid was derived by insertion of the *KpnI/XhoI* fragment of the *LPD* gene into the polylinker region of pUC18. ▬, Insert DNA containing the upstream region and 188 bp of the coding region of the *LPD* gene. Restriction sites are: *Kpn*I (K); *Sal*I (S); *Sst*II (St); *Xba*I (X).

2.2 CULTURE CONDITIONS AND GROWTH MEDIA

2.2.1 Cultivation of *E. coli*

Strains of *E. coli* were routinely maintained on LB-agar plates (with antibiotic supplementation for maintenance of plasmids), and subcultured every 2-3 weeks. Permanent stocks were made as 20% (v/v) glycerol cultures using stationary phase liquid cultures, stored at -70°C. Cultures were grown at 37°C with shaking at 180 rev min^{-1} on an orbital shaker.

Media used are listed below. Amounts given are per litre:

LB-Medium: Bactotryptone 10 g; NaCl 5 g; Yeast Extract 5 g.

M9 salts (10x): Na_2HPO4 60 g; KH_2PO_4 30 g; NaCl 5 g; NH_4Cl 10 g.

M9 Medium: 10x M9 salts 100 ml; H_2O 887 ml. The M9 salts solution was autoclaved and allowed to cool before the following were added: 1M $MgSO_4$ 2 ml; 1M glucose 10 ml; 1% thiamine 1 ml. These three solutions were sterilized separately by filtration (thiamine) or autoclaving.

Media containing agar for plates was made according to the appropriate formula for liquid media with the addition of 20 g l^{-1} Oxoid No 3 agar. After cooling to 55°C the vitamin or antibiotic supplements were added and the plates poured.

2x YT Top Agar. Per litre: Bactotryptone 8 g, Yeast Extract 5 g, NaCl 5 g, Oxoid No 3 agar 7 g.

Antibiotics. For the selection and maintenance of ampicillin resistant transformants media were supplemented with 50 μg ml^{-1} ampicillin.

2.2.2 Cultivation of *Saccharomyces cerevisiae*.

Yeast strains were maintained on YEPD plates at 4°C. Permanent stocks were made by suspending stationary phase liquid cultures in 20% (v/v) glycerol and stored frozen at -70°C. Cultures were grown at 30°C and 180 rev min^{-1} (usually as 500 ml cultures in 2 l Erlenmeyer flasks) on an orbital shaker.

Media used are listed below. Amounts given are per litre:

SD Minimal Medium: yeast nitrogen base (w/o amino acids & ammonium sulphate) 1.7 g; D-glucose 20 g; ammonium sulphate 5 g; agar 15 g (for solid media). 3-amino-1,2,4-triazole (3AT) was added to a final concentration of 10 mM to induce the general amino acid response.

YEPD Medium: D-glucose 20 g; peptone 20 g; yeast extract 10 g; agar 20 g (for solid media).

YEPG Medium: Glycerol 30 ml; peptone 20 g; yeast extract 10 g.

2.3 TRANSFORMATION OF *E. coli*

E. coli strains were made competent according to the method of Dagert and Ehrlich (1979), and stored at -70°C. Competent cells were transformed as described by Messing (1983).

2.4 DNA ISOLATION AND PURIFICATION

Small-scale preparations of plasmid DNA from *E. coli* were obtained using the alkaline SDS method of Birnboim & Doly, (1975). Large scale preparations were obtained in the same manner followed by centrifugation in CsCl-ethidium bromide gradients (Maniatis *et al.*, 1982).

DNA concentrations were estimated from the absorbance of the solution at 256 nm. Calculations were based on an OD_{256} = 1 being equivalent to 50 $\mu g\ ml^{-1}$.

2.5 GENERAL RECOMBINANT DNA TECHNIQUES

2.5.1 Restriction Digests of DNA

Restriction enzymes were used as recommended by the supplier (B. R. L., N. B. L. and Boehringer).

2.5.2 Ligation of DNA

Ligation was carried out using T4 DNA ligase following the protocol of Maniatis *et al.* (1982).

2.5.3 Agarose Gel Electrophoresis

DNA fragments were separated by horizontal, gel electrophoresis using the Tris/borate buffer system described in Maniatis *et al.* (1982).

2.5.4 Electro-elution of DNA

Digested fragments of DNA were recovered from agarose gels by electroelution onto dialysis membrane (Smith, 1980), or using a biotrap chamber according to the manufacturer's instructions (Schleicher & Schuell).

2.5.5 Radio-active labelling of DNA

DNA fragments were labelled with [α-^{32}P]dATP or [α-^{32}P]dCTP (Amersham) either by nick translation using the BRL kit, end-labelling as decribed by Maniatis *et al.* (1982), or by oligo-primed labelling using the method of Feinberg & Vogelstein (1984).

2.6 CLONING INTO THE M13-BASED VECTOR

Shotgun cloning into M13-based vectors was carried out according to Messing, (1983). Transfection of *E. coli* JM101 was performed according to Winter & Fields (1980).

2.7 NUCLEOTIDE SEQUENCE ANALYSIS

Single-stranded M13 DNA templates were prepared and sequenced by the dideoxy chain termination method using a 17-nucleotide synthetic primer (Sanger *et al.*, 1977). Two 15-nucleotide synthetic primers, based on information obtained from the *LPD* sequence, were also used. The nucleotide sequence was compiled and analysed using the University of Wisconsin Genetics Computer Group programmes (Devereux *et al.*, 1984).

2.8 METHODS FOR ANALYSIS OF RNA

2.8.1 Isolation of RNA from *Saccharomyces cerevisiae*

Total yeast cellular RNA was prepared by the method of Weir-Thompson & Dawes (1984).

PolyA$^+$-RNA was isolated according to Roy and Dawes (1987).

2.8.2 Gel Electrophoresis of RNA

RNA fractions were denatured with formaldehyde then separated by electrophoresis through 1% agarose gels containing 2.2 M formaldehyde as described by Maniatis *et al.* (1982).

2.8.3 Northern Hybridisation

Following electrophoresis, the RNA was treated with alkali and neutralised according to Maniatis *et al.* (1982). The RNA fractions were transfered to nylon filters (Amersham). The nylon filters were fixed by UV-irradiation and hybridised with labelled DNA probes following the protocol of the supplier. The hybridisation filters were exposed to Agfa-Gevaert, Curix RP1 X-ray film at -70°C with intensifying screens. Films were developed using Kodak LX-24 and fixed with Kodak FX-40.

2.8.4 Slot blots

RNA was concentrated onto nylon filters essentially according to the method of Tltsty (1982). 15 µl 20 x SSC and 10 µl 37% (w/v) formaldehyde were added to 5 µg polyA$^+$-RNA and made up to a final volume of 50 µl. Samples were denatured by incubating them at 60°C for 15 min. 25 µl were transfered to nylon using BRL's 'Hybri-slot' apparatus. Filters were then treated as described above (2.8.3). Densitometry was performed by scanning autoradiographs using Shimadzu Dual-Wavelength Chromato Scanner CS-930 apparatus.

2.9 IN VITRO TRANSCRIPTION AND TRANSLATION

The expression vector pSP64-GCN4 was purified by equilibrium centrifugation in CsCl, linearized with *Eco*RI and transcribed using the method of Krieg and Melton (1984). 0.5 µg of linearised pSP64-GCN4 was transcribed in 40 mM Tris, pH 7.5, 6 mM $MgCl_2$, 10 mM

dithiothreitol, RNAsin (1 unit/µl), 100 µg/ml BSA, 500 µM of each rNTP (except GTP which was added to give a final concentration of 50 µM) and 500 µM diguanosine triphosphate (G-5'ppp5'-G). 1 unit of SP6 RNA polymerase was added and the final volume made up to 10 µl with distilled water. The transcription mix was then incubated at 40 °C for 10 minutes after which GTP was added to 500 µM and the mix incubated for a further 50 minutes at 40 °C. Such a reaction usually generated roughly 1 µg of total RNA, as estimated using agarose gel electrophoresis as used for analysis of DNA. The RNA was extracted with phenol and precipitated with ethanol.

In vitro translation was carried out using a wheat germ extract as directed by the manufacturers (Bethesda Research Laboratories). Approximately 100 ng of *in vitro* synthesised mRNA was added to a final volume of 30 µl containing potassium acetate, pH 7.5, 33 mM, magnesium acetate 0.6 mM, 16 µCi ^{35}S-methionine, 10 µl nuclease treated wheat germ extract and 3 µl of a reaction mixture containing all of the amino acids with the exception of methionine. Translation products were examined by electrophoresis in a 15 % polyacrylamide gel (30 parts acrylamide: 0.8 parts bisacrylamide) containing SDS as described by Laemmli (1970). After electrophoresis the gel was fixed and autoradiographed.

2.10 ISOLATION OF DNA-BINDING ACTIVITY

DNA-binding proteins were isolated from *Saccharomyces cerevisiae* using a scaled down version of the method of Ruet *et al.* (1984). Yeast cells were grown in 10 l of YEPG and collected in the exponential phase of growth by centrifugation at 2000 x g. The harvested cells were washed in 1 pellet volume of ice cold buffer C (200 mM Tris-HCl pH 8; 10 mM $MgCl_2$ $6H_2O$; 10 mM 2-mercaptoethanol; 1 mM Na_2EDTA; 1 mM PMSF; 10 % v/v glycerol) containing 0.3 M $(NH_4)_2SO_4$. The usual yield was approximately 25 g wet weight. The cell pellet was resuspended in buffer C, containing 0.3 M $(NH_4)_2SO_4$, at 2 g ml^{-1} and transfered to a Braun homogeniser breakage vial on ice. Acid washed 40-mesh glass beads were added until the liquid level above the beads was 5 mm and immediately before breakage PMSF was added to a final concentration of 1 mM. Cells were broken by agitation in a Braun Homogeniser (3 x 1 min bursts with 5 min periods of cooling between). Glass beads were removed by low speed centrifugation (2000 x g, 2 min) through a stainless steel mesh screen. The yeast extract was centrifuged at 40,000 g for 1 hr at 4°C. The supernatant was removed with a Pasteur pipette taking care to avoid the overlying lipid layer and the soft upper region of the pellet. The supernatant was diluted with an equal volume of column buffer (20 mM Tris-HCl pH 8; 1 mM Na_2EDTA; 10 mM 2-mercaptoethanol, pH 8.0) and applied to a 50 ml heparin-Sepharose column equilibrated with column buffer containing 0.1 mM $(NH_4)_2SO_4$ at a flow rate of 1 ml min^{-1}. The column was washed with column buffer containing 0.1 M $(NH_4)_2SO_4$ until unbound protein had eluted. Protein fractions (4 ml) were eluted by a 200 ml salt gradient from

0.1 to 0.75 M $(NH_4)_2SO_4$ and were dialysed against storage buffer (20 mM Tris-HCl pH 8; 50 mM KCl; 1 mM dithiothreitol; 0.2 mM EDTA; 10 % v/v glycerol) overnight at 0 - 4°C and then stored at -70°C in small aliquots. The protein concentration of the fractions was measured using the Biorad assay according to the manufacturer's instructions.

2.11 ANALYSIS OF PROTEIN-DNA COMPLEXES BY GEL RETARDATION

2.11.1 Gel retardation using radio-labelled GCN4

The DNA-binding assay used to investigate the complexes formed between radio-labelled GCN4 and DNA fragments from the *LPD* gene was based upon the procedures described by Garner and Revzin (1981) and Fried and Crothers (1981). Binding was carried out in 15 µl of 20 mM Tris (pH 7.4), 50 mM KCl, 3 mM $MgCl_2$, 1 mM EDTA, 100 µg ml^{-1} gelatin, and 50 µg ml^{-1} sonicated salmon sperm DNA (binding buffer). The DNA to be bound was included to a final concentration of approximately 10 nM. Translation products (0.5 µl) from the *in vitro* translation of GCN4 were added without further purification and the mixture incubated for 20 min at 25°C. 5 µl of loading buffer (binding buffer containing 20 % glycerol, 1 mg ml^{-1} xylene cyanol FF, and 1 mg ml^{-1} bromophenol blue) was added to the reaction mix. The samples were then immediately loaded on to a 5% polyacrylamide gel (30 parts acrylamide: 0.8 parts bisacrylamide) and electrophoresed in 90 mM Tris-Borate buffer (pH 8.3), which was recirculated, at 400 volts until the samples entered the gel and then at 175 volts until the bromophenol blue had migrated the

length of the gel (20 cm). Gels were run at 4°C. The gel was then fixed and autoradiographed.

2.11.2 Gel retardation using radio-labelled DNA

The DNA-binding activity of protein fractions was assayed using a method based upon that described by Huet *et al.* (1985), with modifications. Typically, 10 - 20 ng of end-labelled DNA was added to 10 µg of poly(dI):poly(dC) or 10 µg of DNA from a pBR322 plasmid containing an insert of *Deinococcus radiodurans* DNA (gift of Ian Masters Department of Microbiology, University of Edinburgh) and made up to 100 µl with TE buffer (10 mM Tris HCl pH 8.0, 1 mM Na_2EDTA). The solution of labelled and carrier DNA was then extracted once with 100 µl $CHCl_3$ and then ethanol precipitated. The precipitated DNA was collected by centrifugation in a benchtop microcentrifuge for 20 minutes at 4°C. The supernatant was removed and the pellet dried *in vacuo*. The DNA pellet was dissolved in 85.8 µl distilled water and then 26 µl 10X footprint buffer (0.2 M TrisHCl, pH 8.0, 0.7 M KCl, 50 mM $MgCl_2$, 5 mM $CaCl_2$, 5 mM dithiothreitol, 1 mM Na_2EDTA) and 18.2 µl glycerol were added and mixed thoroughly. For a 20 sample assay 6 µl of the above '2X footprint mix' was incubated with varying amounts of heparin-Sepharose fractions (usually 0.3 to 0.6 µg of protein) and distilled water to make the volume up to 12 µl which was incubated at 25°C for 20 minutes and then loaded immediately on to a 5% (w/v) polyacrylamide gel. Electrophoresis was carried out at 400 volts until the samples had entered the gel and then at 250 volts at 4°C with buffer circulation. Loading buffer (20% w/v glycerol, 1 mg ml^{-1} xylene cyanol FF, 1 mg ml^{-1} bromophenol blue) was normally

added to lanes 1 and 20. The time the gel was run for depended upon the size of the radiolabelled DNA fragments used. Fragments of 50 bp migrated at approximately the same rate as the bromophenol blue dye and those of 450 bp with the xylene cyanol dye. Following electropheresis the gels were dried and autoradiographed at -70°C using intensifying screens.

The 23 nucleotide oligonucleotide used in the competition experiments was obtained from the OSWEL DNA synthesis service and annealed and ligated based on the method of Kadonaga & Tjian (1986). 440 µg of the 23mer was dissolved in 67 mM TrisHCl buffer (pH 7.6) containing 13 mM $MgCl_2$, 6.7 mM dithiothreitol, 1.3 mM spermidine and 1.3 mM EDTA in a total volume of 75 µl. This mixture was incubated at 88°C for 2 min, 65°C for 10 min, 37°C for 10 min and room temperature for 5 min. Annealed oligonucleotides were observed by agarose gel electrophoresis and compared with single stranded oligonucleotides prepared by placing the original untreated oligonucleotide in a boiling water bath prior to loading on the agarose gel. It was noted that the majority of the oligonucleotide received from the OSWEL DNA synthesis service was in the double stranded form. For ligation the annealed oligonucleotide was precipitated with ethanol, dried *in vacuo* and then dissolved in 88 mM TrisHCl, pH 7.5 containing 13.3 mM $MgCl_2$, 20 mM dithiothreitol and 1.3 mM spermidine in a total volume of 75 µl and the ligation reaction was initiated by the addition of 20 µl of 20 mM ATP, pH 7 and 5 µl of T4 DNA ligase (10-30 Weiss units) to give a final volume of 100 µl. This mixture was incubated at 16°C for 4 hr and the DNA then phenol-extracted, precipitated with

ethanol, dried *in vacuo* and dissolved in 100 µl water. Ligation of the annealed oligonucleotide was observed as a smear of DNA which migrated slower than the unligated annealed oligonucleotide during agarose gel electrophoresis.

2.12 DNase I FOOTPRINT ANALYSIS

The DNase I footprinting technique used was based upon those described by Galas and Schmitz (1978) and Hope & Struhl, (1985). A 172 bp *Mae*III/*Sst*II fragment isolated from the upstream region of the *LPD* gene was end-labelled to produce an asymmetrically labelled fragment. Approximately 1 ng of this labelled DNA fragment was dissolved in 15 µl binding buffer (20 mM Tris, pH 7.4, 50 mM KCl, 3 mM $MgCl_2$, 1 mM EDTA, 100 $\mu g\ ml^{-1}$ gelatin, and 50 $\mu g\ ml^{-1}$ sonicated salmon sperm DNA), combined with various quantities of the GCN4 translation product (without further purification) and made up to 20 µl with distilled water. After equilibration for 20 min at 25°C, 5 µl of DNase I (20U μl^{-1}) (Pharmacia) was added. DNase I digestion was terminated after 1 min at 25°C by the addition of 25 µl of 0.5% SDS, 10 mM EDTA, 0.5 mg ml^{-1} *E. coli* tRNA. The DNA was extracted with phenol, precipitated with ethanol, redissolved in 80% formamide, and analyzed as described for DNA sequencing. To relate the observed bands with positions of DNase I cleavage a dideoxy sequencing ladder was run alongside the cleavage products.

CHAPTER 3 SEQUENCE ANALYSIS OF THE *LPD* GENE

3.1 NUCLEOTIDE SEQUENCE OF THE *LPD* GENE

The strategy adopted for sequencing the *LPD* gene was based upon the consistent presence of a 3.7 kb *Xho*I fragment within 12 independently isolated cloned sequences, which complemented the *lpd1* mutation (Dickinson *et al.*, 1986). In addition, deletion analysis of pGP1, a YEp13-based plasmid containing a 5.5 kb insert carrying the *LPD* gene, confirmed this 3.7 kb region contained the *LPD* gene (Roy and Dawes, 1987). The *Xho*I fragment was isolated from pGP-R1, a derivative of pGP1 containing no yeast DNA except the 5.5 kb insert. A region of 2.7 kb was sequenced (Ross *et al.*, 1988). All of the sequence was obtained from at least two independent clones and was, with the exception of a 0.1 kb region 0.8 kb upstream of the coding region, derived from both DNA strands. A summary of the sequence data obtained is shown in figure 3.1.

The programme MAP was used to identify the region representing the *LPD* gene. One large open reading frame of 1.5 kb was found, which is consistent with the known size of the yeast lipoamide dehydrogenase subunit polypeptide. The open reading frame exhibits a significantly high score with respect to prefered codon usage in *S. cerevisiae* determined using the programmes CODONFREQUENCY and CODONPREFERENCE. This is shown in figure 3.2. The complete and unambiguous sequence of the *LPD* gene and the primary structure of lipoamide dehydrogenase, translated from the DNA sequence, are shown in figure 3.3.

The nucleotide sequence for the gene encoding lipoamide dehydrogenase in S. cerevisiae has recently been reported separately

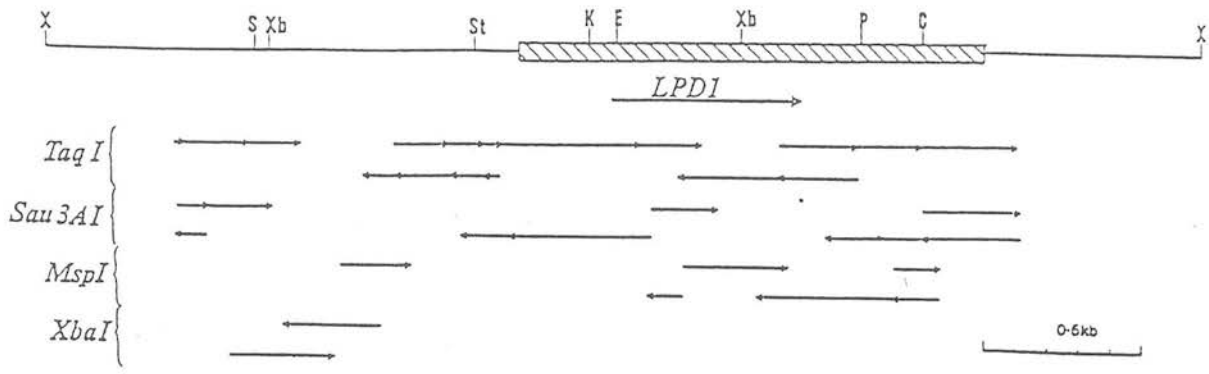
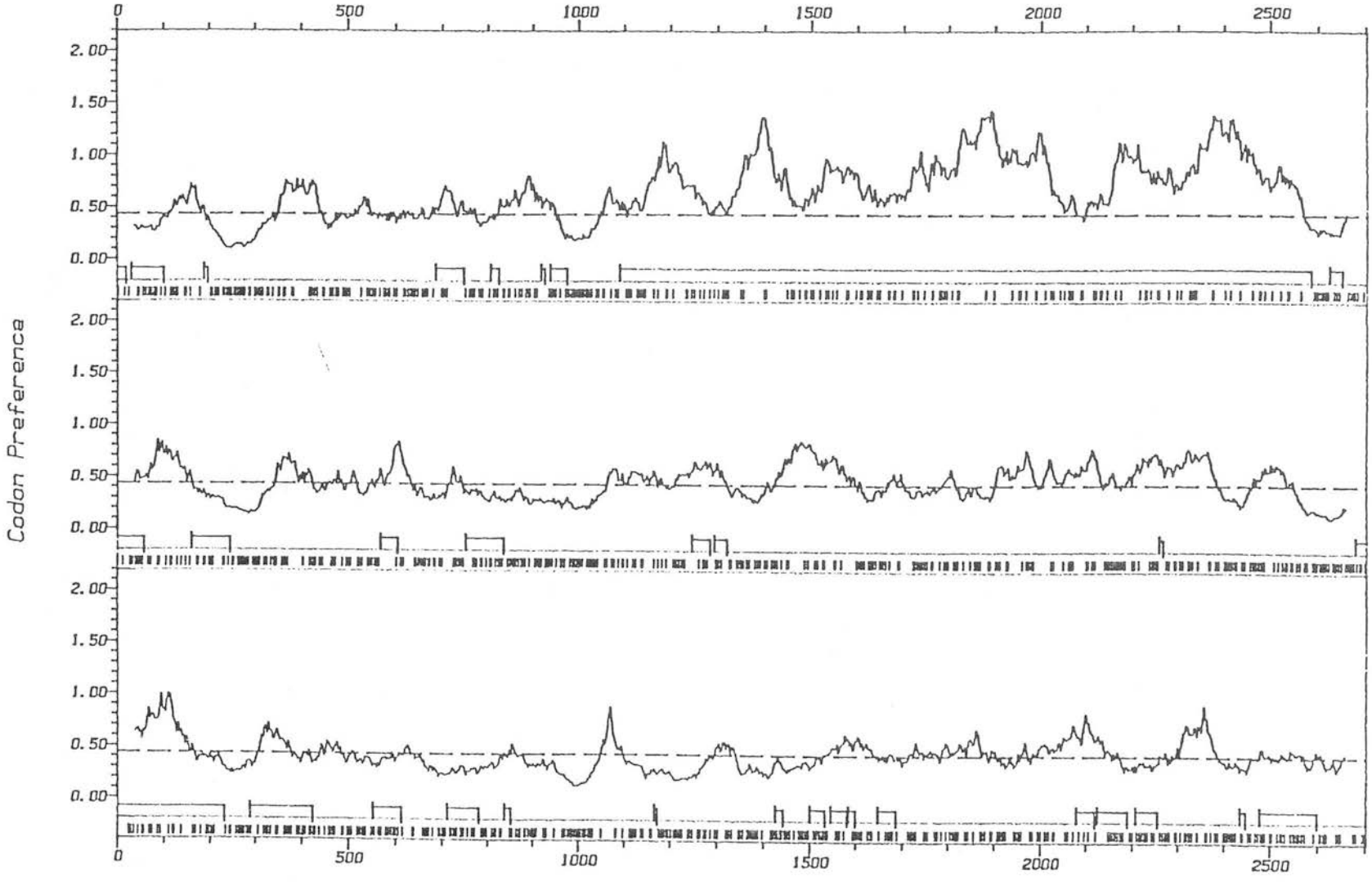


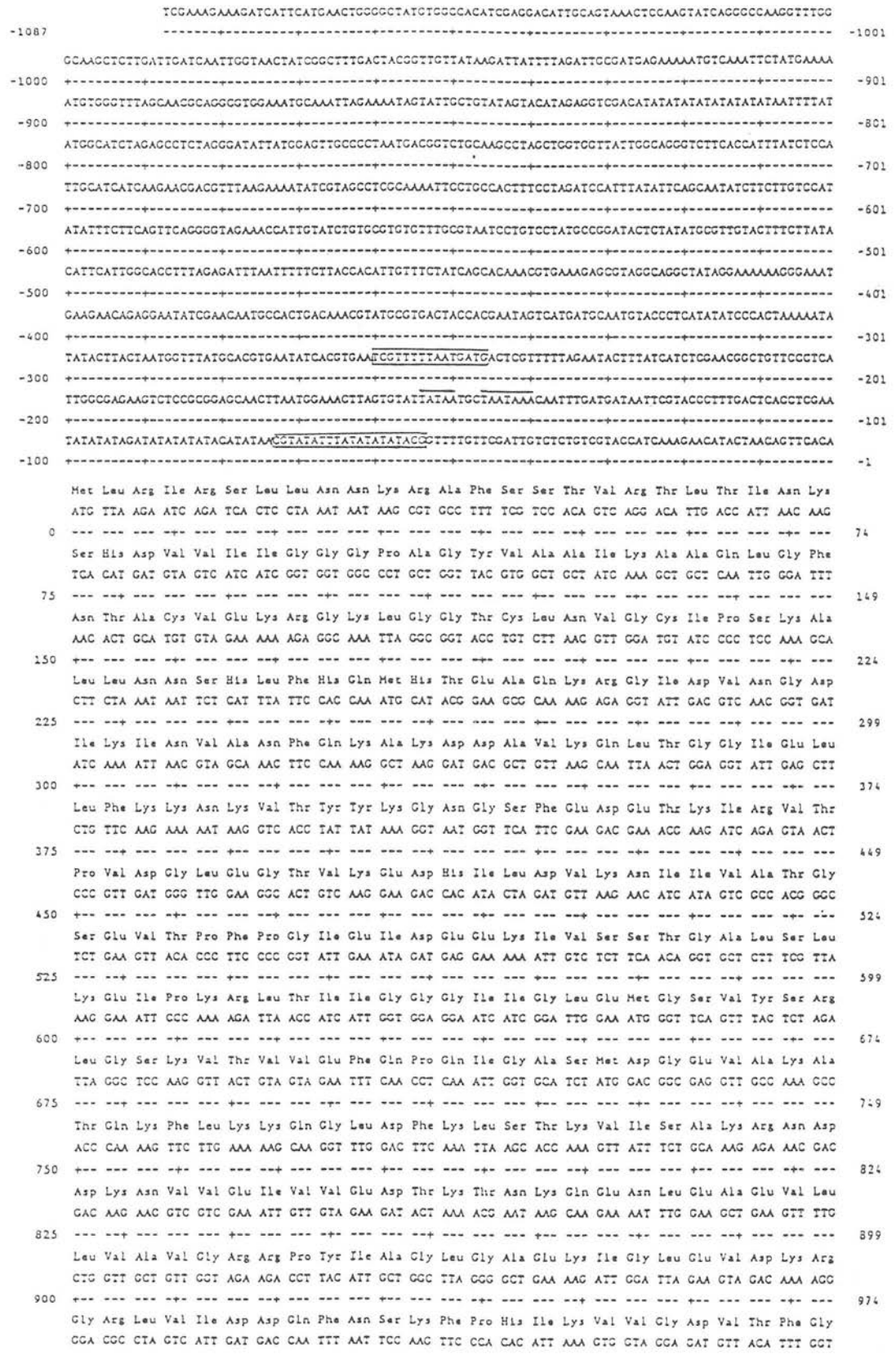
Figure 3. 1. Restriction map of the *LPD* gene and summary of the sequence information obtained

Restriction sites are indicated for enzymes recognising 6 bp sequences (X, *Xho*I; S, *Sal*I; Xb, *Xba*I; St, *Sst*II; K, *Kpn*I; E, *Eco*RV; P, *Pvu*II; C, *Cla*I). The cross hatched region indicates the coding sequence of the gene. The direction of transcription is from left to right. The arrows show the position and extent of DNA sequence obtained from the M13 clones. The enzyme used to digest the 3.7 kb fragment to obtain each group of M13 clones is indicated at the side of the arrows. The sequence is fully overlapping and most of it was obtained from both DNA strands.

Figure 3. 2. Codon preference plot of the 2.7 kb sequenced region.

The codon preference plots of the three forward frames of the 2.7 kb region of DNA sequenced are shown. The programme CODONFREQUENCY was used to construct a codon usage table based upon several highly expressed *S. cerevisiae* genes. The table was then used to search for regions showing a codon utilization bias towards a particular family of codons using the programme CODONPREFERENCE. The broken line at 0.45 is the threshold above which the codon usage is regarded as significantly biased. The open bars below the plot represent open reading frames. The vertical lines, below the open reading frame markers, indicate rare codons. The 1.5 kb *LPD* open reading frame can be seen in frame A and corresponds with a biased codon usage and a reduction in the frequency of rare codons.


Codon Preference
0
500
1000
1500
2000
2500
2.00
1.50
1.00
0.50
0.00



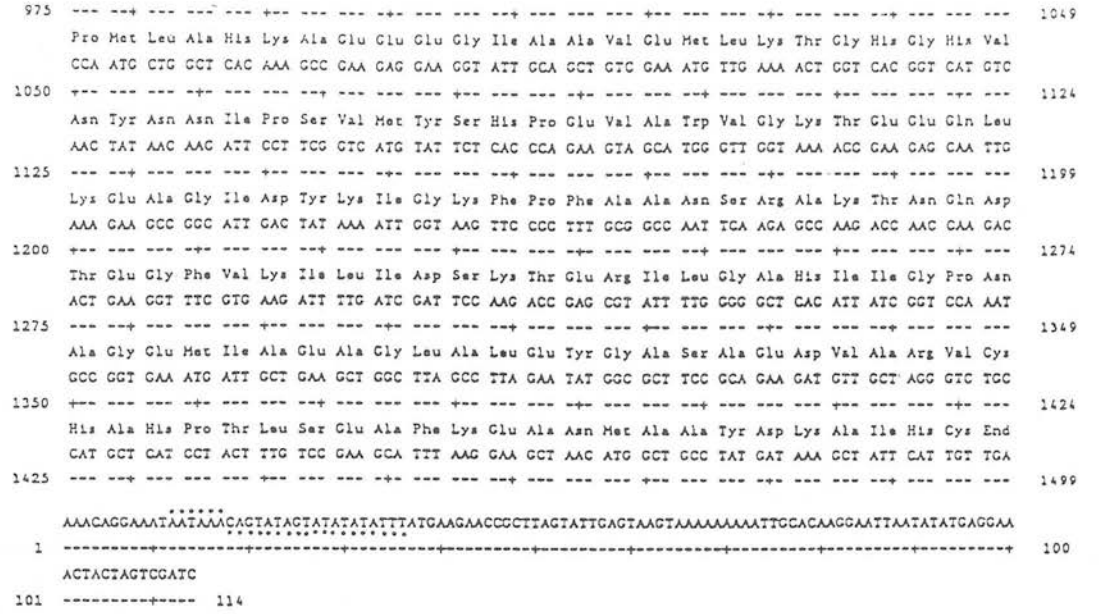


Figure 3. 3. Nucleotide sequence of the *LPD* gene and deduced primary structure of lipoamide dehydrogenase.

The nucleotide sequence of a 2.7 kb region containing the sense strand of the *LPD* gene as well as over 1 kb of flanking DNA is shown in the 5' to 3' direction. The specific features of the 5' and 3' flanking regions are noted as follows: the 'TATA' consensus sequences are overlined; sites showing homology to sequences known to be important in the structure of *ARS* elements, are boxed; the proposed consensus polyadenylation signal, AATAAA, is denoted by overdots; the proposed polyadenylation/termination signal, TAG...TAGTGT...TTT, is denoted by underdots. The sequences showing homology to known regulatory sequences are shown separately in figure 3. 6

(Browning *et al.*, 1988). This nucleotide sequence was determined using clone, isolated from a cDNA library screened with a 17 nucleotide mixed oligonucleotide probe. The probe was based on the amino-terminal sequence of the yeast lipoamide dehydrogenase. The two sequences are identical, with the exception of a one base mismatch at position +282, within the open reading frame which is in the third non-conserved base of a codon. The 5' noncoding regions, however, show no significant homology. The cDNA-derived sequence shows very strong homology to the *ENO1* gene from *S. cerevisiae* and two cDNA molecules must have become ligated during the cloning procedure.

In the region 5'-distal to the *LPD* gene there are two potential open reading frames: one (at -571), read in the opposite sense, is capable of encoding a polypeptide of 77 residues; the other (sequence not complete, terminating at -858) could encode a polypeptide of at least 76 amino acids. A search of the NBRF protein data base showed no strong homology between the two predicted amino acid sequences and that of any known protein.

3.2 PRIMARY STRUCTURE OF YEAST LIPOAMIDE DEHYDROGENASE

The primary structure translated from the nucleotide sequence of the *LPD* gene, contains 499 amino acid residues that correspond to a protein of M_r 54,010 (54,730 including the FAD cofactor). These M_r values are in good agreement with previous estimates (Wieland, 1983). Lipoamide dehydrogenase, encoded in the nucleus and synthesised on cytoplasmic ribosomes, is a component of two multienzyme complexes which function in the mitochondrial matrix-

inner membrane compartment. The presence of an N-terminal presequence has been shown in mammalian cells. Antibodies raised against the E1, E2 and E3 components of 2-oxoglutarate dehydrogenase from ox heart have been used to show the presence, in cultured pig kidney cells, of initial cytoplasmic translation products that are larger than the mature proteins. The E3 component was shown to be synthesised as a polypeptide 10-20 amino acids larger than the mature protein (Hunter and Lindsay, 1986).

The primary structure, derived from the *LPD* nucleotide sequence of *S. cerevisiae*, supports the conclusion that, in this organism, the lipoamide dehydrogenase N-terminal region represents a mitochondrial targeting sequence. The first twenty amino acids are rich in seryl, threonyl and basic residues, and show a complete absence of acidic amino acids. All of these features are typical of mitochondrial targeting sequences in yeast (von Heijne, 1986). In *E. coli* such a targeting sequence would be unnecessary; and when the primary structures of *E. coli* and *S. cerevisiae* (Stephens *et al.*, 1983) lipoamide dehydrogenase are aligned the yeast sequence is seen to contain an additional 20 amino acids at its N-terminus which are absent from the inferred *E. coli* protein (see figure 3.4).

The amino-terminal sequence of the native lipoamide dehydrogenase has recently been determined (Browning *et al.*, 1988). It has the sequence Xaa-Ile-Asn-Lys-Ser-His-Asp-Val-Val-Ile-Ile-Gly-Gly-Gly-Pro-Ala-Gly-Tyr-Val-Ala-Ala-Ile-Lys-Ala-Ala. This confirms that the first 21 amino acids of the nascent lipoamide dehydrogenase polypeptide represent a presequence, probably involved in directing the protein to its correct subcellular location and

facilitating its entry into the mitochondrion. This is cleaved to produce the mature protein.

3.2.1 Comparison of related amino acid sequences

The primary structures of several flavoproteins have been determined either partially or completely. Pig heart lipoamide dehydrogenase, *E. coli* lipoamide dehydrogenase, human red blood cell glutathione reductase and *Psuedomonas aeriginosa* transposon mecuric reductase amino acid sequences have previously been compared (Williams *et al.*, 1984). Four regions were identified which show strong homology between the four proteins. Two of these (FAD-1 and FAD-2) are involved in the binding of FAD. The other two constitute a pyridine nucleotide-binding domain and an 'interface' domain involved in the interaction of subunits of the enzyme producing the dimeric lipoamide dehydrogenase structure.

The primary structure of lipoamide dehydrogenase from *Azotobacter vinlandii* (Westphal and Kok, 1988), and human liver cells (Pons *et al.*, 1988) have recently been deduced from their respective DNA sequences. Figure 3.4 shows the amino acid sequences of the four proteins, previously compared, and the primary structures of lipoamide dehydrogenase from *S. cerevisiae*, *Azotobacter vinlandii* and human liver cells, aligned for maximal homology using the computer programmes GAP and PRETTY.

Table 3.1, constructed using the programme DISTANCES, shows the degree of homology that exists between the 7 different amino acid sequences which have been aligned. The programmes COMPARE and DOTPLOT were used to depict graphically the regions of homology between *S. cerevisiae* lipoamide dehydrogenase and the other six

Figure 3. 4. Protein sequence comparison.

The primary structures of *S. cerevisiae* lipoamide dehydrogenase (YLPDH), human liver lipoamide dehydrogenase (HLPDH), pig heart lipoamide dehydrogenase (PLPDH), *E. coli* lipoamide dehydrogenase (ECLPDH), *Azotobacter vinlandii* lipoamide dehydrogenase (AVLPDH), human red blood cell glutathione reductase (HGR) and *Pseudomonas aeruginosa* mercuric reductase (PAMR) are aligned for maximum homology. The residue numbering is based on *S. cerevisiae* lipoamide dehydrogenase. A consensus line below the seven sequences shows residues common to four or more of the amino acid sequences. The borders of the domains representing the four regions of strong homology previously identified (Williams *et al.*, 1984), are marked by arrowheads below the consensus sequence. The order of the domains is FAD-1, pyridine nucleotide-binding, FAD-2 and finally the interface domain. The 80 residues at the N-terminus of mercuric reductase have been omitted. The individual sequences show residues in agreement with the consensus sequence in upper case and all others in lower case. The alignment was performed using the programmes GAP and PRETTY with the default protein comparison file (Gribskov and Burgess, 1986) which regards certain amino acids as similar for the purposes of the consensus line. The asterisks in the pig heart lipoamide dehydrogenase sequence indicate regions of unknown sequence.


1
57
YLPDH mlrirsllnn krafsstvrt ltinkshDVV IIGgGPaGYV AAIKAAQLGF nTacVEK...
HLPDH .... adqpidaDVt VIGsGPGGYV AAIKAAQLGF kTVcIEK...
PLPDH hfnrishglq glsavplrty adqpidaDVt VIGsGPGGYV AAIKAAQLGF kTVcIEK...
ECLPDH .... .steiktqVV VlGaGPaGYs AAfrcAdLGL eTViVEr...
AVLPDH ....msqkfDVI VIGaGPGGYV AAIKsAQLGL kTalIEKykg
HGR .....acrq epqpqgpppa agavasyDyl VIGgGsGGla sArrAAeLGa raavVE....
PAMR adnrvglldk vrgwmaaaek hsgneppvqV aVigsgGaam AAalkAveqg aqVtlie...
---------- ---------- -------DVV VIG-GPGGYV AAIKAAQLGF -TV--EK---
1 3
...rgkLGGT CLNVGCIPSK ALLNnSHlFH qMhteaqkrg i.dvngDIki NvanfqkaKD
...neTLGGT CLNVGCIPSK ALLNnSHyYH .MahgkdFAs rGiemsEVrl Nlekm*****
...neTLGGT CLNVvCIPSK ALLNnSHyYH .MahgkdFAs rGiemsEVrl NldkmmeqKs
...ynTLGGv CLNVGCIPSK ALL...Hvak vieeakaLAe hGivfgEpkt didkirtwKE
kegktaLGGT CLNVGCIPSK ALLdsSykFH eahesfkLhg i.s.tgEVai dvptmiarKD
...shkLGGT CvNVGCVPkK vMwNtavhse fMh...dhAd yGfpscEgkf NwrvikekrD
...rgTiGGT CvNVGCVPSK iMiraaHiaH lrrespfdgg iaatvptIdr skllaqqqar
-----TLGGT CLNVGCIPSK ALLN-SH--H -M------A- -G----EV-- N-------KD
172
daVkqLTGGI .elLFKknkV tyykGnGsFe detkirVTpv DGlegtVked hildvkNIiV
****aLTGGI ah.LFK**** **VnGyGK** *********a DGstevInt. .....kNIlI
taVkaLTGGI ahlLFKqnkV vhVnGyrKit gknqvtaTka DGgtqvIdt. .....kNIlI
kvInqLTGGl .agMaKgrkV kvVnGlGKFt gantleVege nG.......k tvinfdNaiI
qiVrnLTGGV asl1.KangV tlfeGhGKLl agkkveVTaa DG......ss qvldteNVil
ayVsrLnaiy .qnnLtkshI eiIrGhaaF. ....Tsd pkptieVsgk kytaph.IlI
vdelrhakye .giLggnpaI tvVhGearFk ddqsltVrln EGgervVmfd rcl......V
--V--LTGGI ---LFK---V --V-G-GKF- ------VT-- DG----V--- ------NI-I
231
ATG.SevTPF PgIeIDEekI VsSTGALsLk eIPkRLtIIG gGIIGLEMGs VYsRLGSkVT
ATG.SevTPF PgItIDE*** ********** ****kMvVIG AGVIGvE... **********
ATG.SevTPF PgItIDEdtI VsSTGALsLk kVPekMvVIG AGVIGvELGs VWqRLGadVT
AaG.SrpiqL PfIphEDprI wdSTdALeLk eVPeRLlVmG gGIIGLEMGt VYhaLGSqId
AsG.Skpvei PpapVDqdvI VdSTGALdFq nVPgkLgVIG AGVIvLELGs VWaRLGaeVT
ATGgmpsTPh esqipgaslg ItSdGfFqLe elPgRsvIVG AGyIavEMag ILsaLGSkts
ATG.aspavp PipglkEspy wtSTeALasd tIPeRLaVIG ssVVaLELaq aFaRLGSkVT
ATG-S--TPF P-I-IDE--I V-STGAL-L- -VP-RL-VIG AGVIGLE-G- V--RLGS-VT
288
VvEfqpqIg. asmDgeVaKa tqkfLkKqGl dFklstkVis akrnddknvV eiv..vedtk
********** ********** ********** ********** ********** **********
avEflghVgg vgiDmeIsKn fqriLqKqGf kFklntkVtg atkksd.gkI dv..sieaas
VvEmfdqVi. paaDkdIvKv ftkrisKkfn lMletkvtav eakedgiyv. .....tmegk
VlEamdkf.l pavDeqVaKe aqkiLtKqGl killgarVtg tevknkqvtV kfv....dae
lmirhdkVl. rsfDsmIstn cteeLenaGv evlkfsqVke vkktlsgleV smvtavpgrl
Vlarntlff. redpaigeav taafraegie vLehtqasqv a..... ..hmdgefvl
V-E----V-- ---D--I-K- ----L-K-G- -F-----V-- ---------V ----------
327
tnkqenleaE vLLVAVGRrP yiagLgaEki GlEvDkRGr. ....
********** **LVcIGReP fTqnLgLEel GIElr***** *********. .....
ggkaevitcD vLLVcIGRrP fTknLgLEel GIElDpRGri pvntrfqtki pniyaigdpr
kapaepqryD avLVAIGRvP ngknLdagka GVEvDdRGf. ....
geksqafdk. .LiVAVGRrP vTtdLlaads GVtlDeRG.. ....
pvmtmipdvD cLLwAIGRvP nTkdLsLnkl GIqtDdkGh. ....
ttthgelraD kLLVAtGRtP nTrsLaLDaa GVtvnaqGa. ....
---------D -LLVAIGR-P -T--L-LE-- G-E-D-RG-- ---------- ----------
383
..lvIDdqFn skfPhIkvVG DVtfGPMLAH KAeEEG..IA aVEmLktGhg hvnYNnIPSV
....IPnIaAIG DVVaGPMLAH KAeDEG..Ii cVEgMaGGav hiDYNcVPSV
grIpVntrFq TkIPnIYAIG DVVaGPMLAH KAeDEG..Ii cVEgMaGGav hiDYNcVPSV
..IrVDkqLr TnVPhIFAIG DIVgqPMLAH KgvhEGhvaA eViagk..kh yFDpkvIPSI
.fIyVDdyca TsVPgVYAIG DVVrGaMLAH KAsEEGvvVA ..EriaGhka qMnYdlIPaV
..IiVDefqn TnVkgIYAVG DVcgkaLLtp vAiaaGrklA hrlfeykeds kLDYNnIPtV
..IvIDqgMr TsnPnIYAaG DctdqPqFvy vAaaaG..tr aainMtGGda aLDltamPaV
--I-VD--F- T-VP-IYAIG DVV-GPMLAH KA--EG--IA -VE-M-GG-- -LDYN-IPSV
441
mYsHPEVAWV GKT..EeqLK EaGIDYKIGk FPFAAnsRAk TNqDTEGfVK IliDskTERI
IYTHPEVAWV GKs..EeqLK EEGIEYKVGk FPFA***RAk TNaDTDGmVK IlgqksTDRV
IYTHPEVAWV GKs..EeqLK EEGIEYKVGk FPFAAnsRrk TNaDTDGmVK IlgqksTDRV
aYTePEVAWV GlT..EkeaK EkGIsYetat FPWAAsgRAi asdcaDGmtK lIfDkeshRV
IYTHPEIAgV GKT..EqaLK aEGVainVGv FPFAAsgRAm aanDTaGfVK VIaDakTDRV
VFsHPpIgtV GlTedEaihK ygienvKtys tsFtpmyhAv TkrkTkcvmK mVcankeEkV
VFTdPqVAtV G..ysEaeah hDGIEtdsrt LtLdnvpRAl aNfDTrGfIK lViEegshRl
IYTHPEVAWV GKT--E--LK EEGIEYKVG- FPFAA--RA- TN-DTDG-VK I--D--TDRV
499
LGaHIIGPnA GEMIaEAgLA LEYGASaEDV ArvcHAHPTL SEAFkEAnMA aydkaihc..
LGaHIIGPgA GEMInEAALA LEYGAScEDI ArvcHAHPTL SEAFrEAnLA asfgkainf.
LGaHIlGPgA GgMVnEAALA LEYGAScEDI ArvcHAHPTL SEAFrEAnLA askgksinf*
iGgaIVGtng GELlgEigLA iEmGcdaEDI AltiHAHPTL hEsvglAaev fegsitdlpn
LGvHVIGPsA aELVqqgAiA MEFGtSaEDl gmmvfAHPaL SEALhEAaLA vsghaihvan
vGiHmqGlgc dEMlqgfAvA vkmGAtkaDf dntvaiHPTs SEeLvtlr.. .....
iGvqaVaPeA GELIqtAALA irnrmtvqEl Adqlfpyl TM vEgLklAaqt fnkdvkqlsc
LG-HIIGP-A GEMI-EAALA LEYGAS-EDI A---HAHPTL SEA--EA-LA ----------

	YLPDH	PHLPDH	HLPDH	ECLPDH	AVLPDH	HGR	PAMR
YLPDH	1	0,45	0,62	0,47	0,56	0,40	0,38
PHLPDH		1	0,59	0,37	0,40	0,29	0,28
HLPDH			1	0,48	0,54	0,38	0,36
ECLPDH				1	0,49	0,35	0,37
AVLPDH					1	0,37	0,36
HGR						1	0,32
PAMR							1

Table 3. 1. Score of the homology between the seven amino acid sequences.

The programme DISTANCES was used to score the relative homology between the seven different primary amino acid sequences. Each value represents the degree of homology between two primary sequences determined using the default protein comparison file of the programme (Gribskov and Burgess, 1986). YLPDH, *S. cerevisiae* lipoamide dehydrogenase; HLPDH, human liver lipoamide dehydrogenase; PLPDH, pig heart lipoamide dehydrogenase; ECLPDH, *E. coli* lipoamide dehydrogenase; AVLPDH, *A. vinlandii* lipoamide dehydrogenase; HGR, human red blood cell glutathione reductase; PAMR, *P. aeruginosa* mercuric reductase.

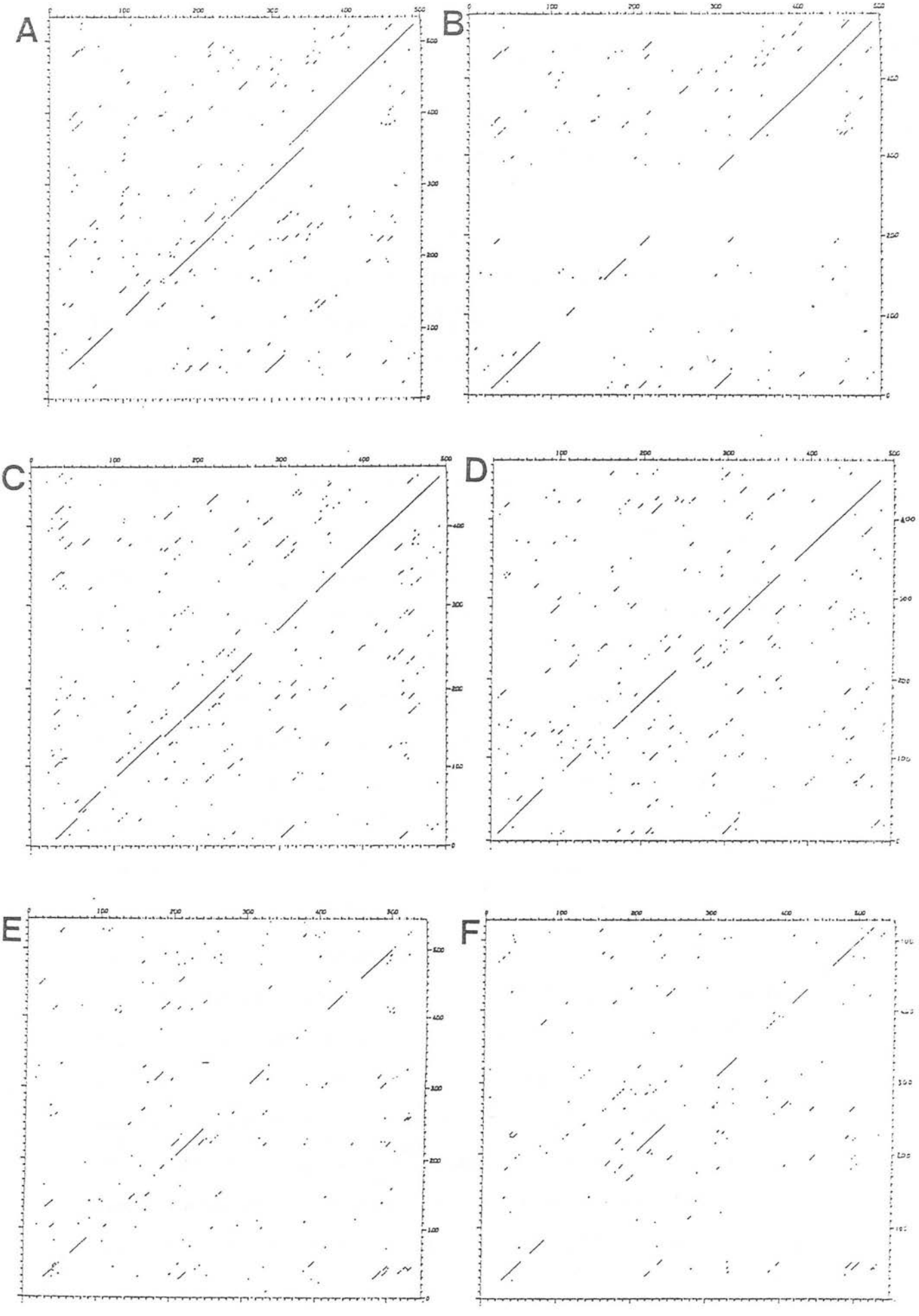
primary sequences. The results of this analysis are shown in figure 3.5.

The primary structure of *S. cerevisiae* lipoamide dehydrogenase is in full agreement with the established 'domain homology' identified between the four previously aligned flavoproteins. The higher level of homology between the eukaryotic amino acid sequences is apparent from figure 3.4 and table 3.1. The relatively low homology between the pig heart sequence and the other primary sequences displayed in the table is due to the regions of unknown sequence in this protein. A glance at the top three lines of figure 3.4 reveals the very high level of conservation present between the yeast and human liver sequences and that of pig heart lipoamide dehydrogenase. Interestingly there is no region of strong homology exclusive to the eukaryotic primary sequences. Such a region might have been expected based on the difference between eukaryotic complex structure and that of Gram-negative prokaryotes and its absence perhaps indicates that the lipoamide dehydrogenase interacts with the E2 subunit in the same manner in both multienzyme complex conformations.

Table 3.1 also reveals the higher degree of homology between *A. vinlandii* and the eukaryotic sequences than that between *E. coli* and these sequences perhaps reflecting a closer relationship, in evolutioary terms, between *A. vinlandii* and eukaryotic organisms. The homology between all five E3 amino acid sequences and the two 'non-lipoamide dehydrogenase' primary sequences is in full agreement with the proposal that lipoamide dehydrogenase represents the progenitor of these two proteins. Williams *et al.*, (1984), have proposed that, as lipoamide dehydrogenase is found in several

Figure 3. 5. Dot plots showing the regions of homology between *S. cerevisiae* lipoamide dehydrogenase and six related amino acid sequences.

In each comparison *S. cerevisiae* lipoamide dehydrogenase is along the horizontal axis and the other primary sequence along the vertical axis. A, human liver lipoamide dehydrogenase; B, pig heart lipoamide dehydrogenase; C, *E. coli* lipoamide dehydrogenase; D, *A. vinlandii* lipoamide dehydrogenase; E, human red blood cell glutathione reductase; F, *P. aeruginosa* mercuric reductase. The numbers along the axes represent the amino acid number for each primary sequence. Regions of homology were identified using the programmes COMPARE and DOTPLOT, with the default threshold value and a window size of 15 residues, and are displayed as diagonal lines between the two sequences.


A
B
C
D
E
F

anaerobic organisms, glutathione reductase may have evolved from this enzyme during the oxygen build-up era as organisms needed to protect their protein thiols against oxidation. Regions of homology can be identified between *S. cerevisiae* and both human glutathione reductase and *P. aeruginosa* transposon mercuric reductase, in figure 3.5, which correspond approximately with all of the four domains previously identified and discussed above.

3.3 GENERAL FEATURES OF THE NONCODING REGIONS

The sequences upstream and downstream which flank the coding region of the *LPD* gene share several characteristics common to the 5'- and 3'-noncoding regions of many sequenced yeast genes. The noncoding regions of several yeast genes are AT rich and the regions 5' (-500 to -1) and 3' to the *LPD* coding region have overall A + T compositions of 63% and 74% respectively. The sequences TAATAAA and TATAA found at positions -146 and -154 represent potential TATA boxes, one of the elements known to have an important role in determining the mRNA initiation site (Grosschedl and Birnsteil, 1980). An A is found at position -3 as has been reported for the majority of yeast genes sequenced (Kozak, 1981).

Two further sites in the upstream region are of interest as they show homology to sequences shown to be important in the structure of autonomously replicating sequences or ARS. First, the region from -73 to -54 with the sequence CGTATATTTATATATATATCG, an almost perfect 21 bp inverted repeat, is similar to a consensus sequence proposed to be necessary for autonomous replication with the sequence $^{A}/_{T}$TTTATPuTTT$^{A}/_{T}$ (Broach *et al.*, 1983). This homology

may be purely fortuitous as the sequence between -73 and -54 lies within a very AT-rich region and fine structure analysis of *ARS* sequences has shown they require a number of additional elements other than the one described above for functional activity (Celniker *et al.*, 1984). Second, a region from -260 to -247 with the sequence TCGTNNNNAATGATG differs by one base from the sequence TCGTNNNNAATGATA which represents the opposite strand of a sequence identified as the consensus recognised by the DNA-binding factor, ABFI (*ARS*-binding factor; Buchman *et al.*, 1988). The search for *ARS*-like sequences was carried out following the observation that an integrating vector containing the upstream region of the *LPD* gene but no yeast origin of replication was maintained to some extent as an independent autonomously replicating plasmid (Z. Zaman unpublished results). ABFI binding sites have been identified in the upstream region of several yeast genes including *HIS3*, *DED1* and *REP3* genes (Buchman *et al.*, 1988). It has been postulated that they perhaps operate as activation sites or as boundary regions separating adjacent genes and they may therefore represent another example of a protein-binding site with more than one function.

Downstream of the TGA translation termination codon an additional 114 bp has been sequenced. The motif AATAAA, postulated to specify a site involved in polyadenylation in higher eukaryotes (Proudfoot and Brownlee, 1976), and seen in many yeast genes, is found at position +12. The sequence TAG...TA(T)GT...(AT rich)...TTT has been postulated to be a signal for transcription termination or polyadenylation (Zaret and Sherman, 1982), and a version of this, CAGTATAGTATATATATTT, is found at position +18.

3.4 POTENTIAL REGULATORY ELEMENTS WITHIN THE *LPD* UPSTREAM REGION

The upstream region of the *LPD* gene, from -358 to -102, contains a number of regions which show homology to sequences known to have roles in the regulation of other yeast genes. These regions of homology are shown in figure 3.6. At position -247 there is the sequence TGACTCGTTTTT that conforms very closely to the consensus sequence obtained from the GCN4 protein binding sites studied (Hill *et al.*, 1986). This protein mediates the general control of amino acid biosynthesis response that is involved in modulating the expression of genes, involved in amino acid biosynthesis, during conditions of amino acid starvation (Struhl, 1982). A second, almost perfect direct repeat, sequence (TGAATCGTTTTT), is also present, 17 bp upstream, at position -265. At -114 there is another TGACTC motif, although this conforms less well to the GCN4 binding site consensus sequence. The role of general amino acid control in the regulation of the *LPD* gene will be discussed more fully in chapter 4.

The sequence at position -203 shows homology to a sequence identified as a component of the UAS2 of *CYC1* (Forsburg and Guarente, 1988). The *CYC1* gene, encoding iso-1-cytochrome *c*, contains two UAS's, termed UAS1 and UAS2, both of which represent examples of complex upstream elements which interact with multiple regulatory factors (Guarente *et al.*, 1984). UAS1 is responsible for the haem-induced activation of *CYC1*. UAS2 is the element which is responsible for the activation of *CYC1* expression in the presence of nonfermentable carbon sources and is activated by the gene products of the *HAP2* and *HAP3* loci. These regulatory genes are part of a control network modulating the expression of genes required for

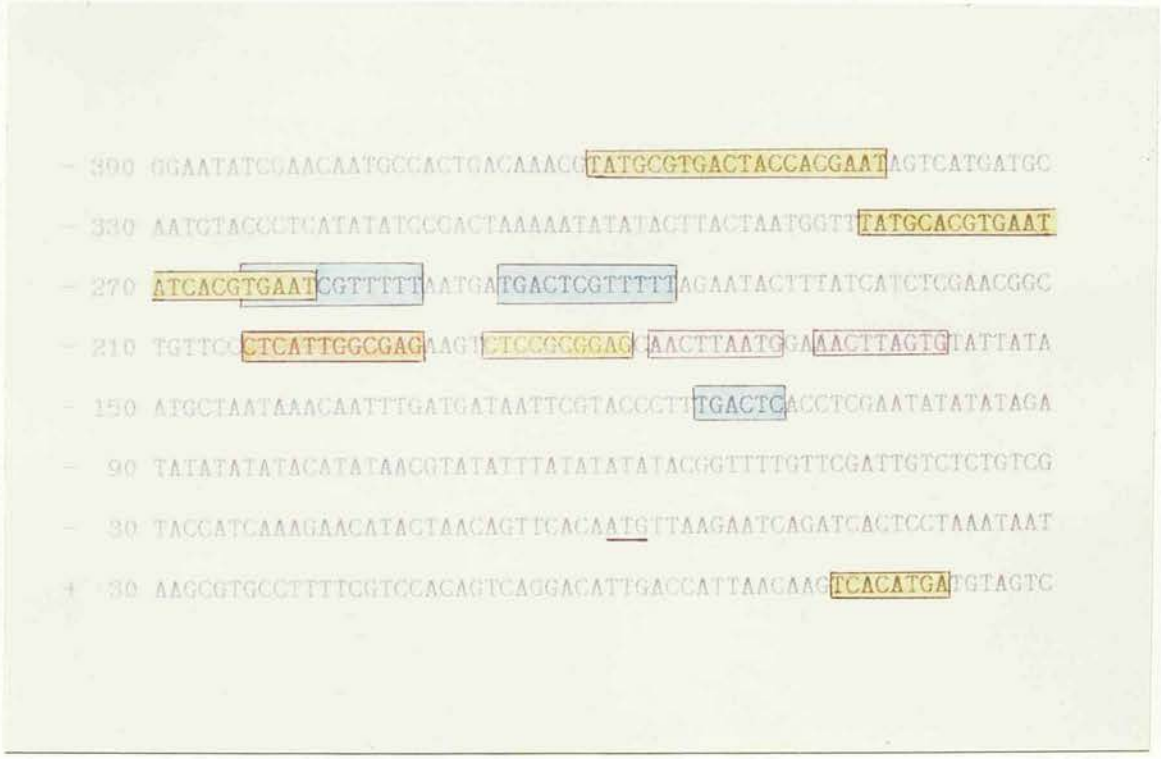


Figure 3. 6. Sequences with a potential role in the regulation of the *LPD* gene.

The boxed regions all show homology to sequences known to have roles in the regulation of other yeast genes. Blue shading, GCN4 binding site homologies; orange shading, homology to UAS2 of *CYC1*; yellow shading, homology to the original 'Cooper repressor sequence'; purple shading, homologies to ADR1 binding site; green shading, homologies to the TCACGTGA sequence present as a *cis*-acting element within several yeast promoters and the core of the CP1 binding site. The sequence from -361 to -342 has also been shaded in green to highlight its similarity to the sequence from -283 to -260. The translation start site is underlined. See text for further details

respiratory competence including *HEM1* and *COX4* (Forsburg and Guarente, 1988).

A detailed mutational analysis of UAS2 involving 5' and 3' deletions and linker substitution analysis has revealed this site contains two distinct regions which bind to specific regulatory factors (Forsburg and Guarente, 1988). Region 1 is the binding site of the HAP2 and HAP3 proteins and consists of the principal carbon source response site. Region 2 binds a distinct protein complex and is required for maximal activity from region 1. The position of these regions within the *CYC1* upstream sequence together with the reported homology identified between *CYC1*, *HEM1* and *COX4*, are shown in figure 3.7. The consensus sequence TNATTGGT has been identifed between these three genes based upon a region of homology between them and the observation that a one base change within the UAS2 of *CYC1* at position -208, shown in the figure, increases the activity of the site 10 to 20-fold (Guarente *et al.*, 1984). Within region 1 a dyad sequence is also present and has been shown to be required for maximal UAS2 activity. The regions of homology, between the *LPD* gene and those identified as the *HAP2-HAP3* binding site, are shown in figure 3.7. No further homology has been reported between the regulatory regions *CYC1*, *HEM1* and *COX4* genes.

The sequence of the *LPD* gene showing homology to the identified consensus sequence is also shown in figure 3.7. The addition of the *LPD* sequence helps demonstrate two further similarities that exist between all 4 sequences First, a sequence of dyad symmetry, in the region immediately downstream of the consensus sequence, present in all four sequences. This sequence, within the *CYC1*, *HEM1* and *LPD* genes, shows some degree of homology;

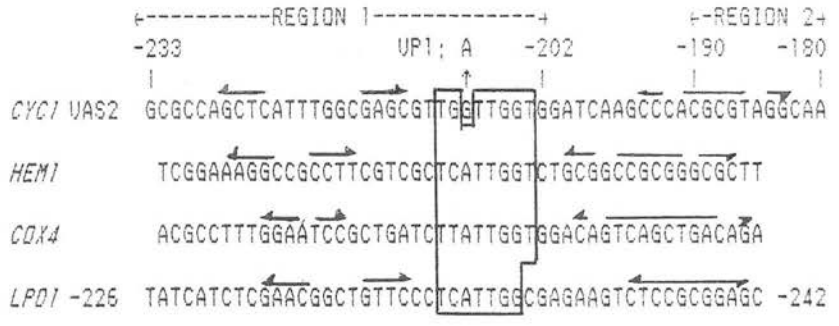


Figure 3. 7. Sequences showing homology to the HAP2-HAP3 binding site within the UAS2 of *CYC1*.

The consensus sequence, previously identified (Forsburg and Guarente, 1988), to which the *LPD* gene shows homology, is boxed. Regions 1 and 2 define two separate sites identified as the binding sites of regulatory factors. The inverted repeats, discussed in the text are shown above each DNA sequence. UP1 refers to the mutation in the *CYC1* promoter region which increases UAS2 activity (see text for further details).

all three contain the sequence CGCG which extends to CCGCGG in *HEM1* and *LPD*. Within the UAS2 of *CYC1* this region of dyad symmetry may represent the binding site of the protein, identified by gel retardation studies as binding to region 2. Second, all four sequences contain short regions of dyad symmetry similar in nature but not in sequence to that identified as an important element of the *CYC1* UAS2. Such a region of dyad symmetry may be necessary for the binding of the HAP2 and HAP3 proteins. The location of these two further features is also shown in figure 3.7.

The presence of a sequence similar to the UAS of *CYC1* within the *LPD* promoter is not suprising given the role this *cis*-acting element has in mediating the catabolite repressible nature of iso-1-cytochrome *c*. *LPD* expression has been shown to be subject to catabolite repression (Roy and Dawes, 1987),and a UAS2-like element should confer this form of regulation upon the gene. It is interesting that both *CYC1* and *LPD* show a one base mismatch to the TNATTGGT consensus sequence. Clearly, the presence of an exact copy of this sequence is not required for a functional HAP2-HAP3 binding site. How important the different bases of this sequence and the sequences surrounding it are remains to be determined. Results connected with the presence of a HAP2-HAP3 binding site in the upstream sequence of the *LPD* gene together with proposals for the further investigation of this possibility are discussed in Chapter 5.

The inverted repeat downstream of the HAP2-HAP3 binding site is also interesting as it shows some homology to the sequence TAGCCGCCGAGGG previously proposed by T. Cooper (1986) to represent a general repressor protein binding site. Since the original

proposal, however, the exact sequence of this binding site has been re-examined and revised. The revised sequence is not yet available, although it is known not to be a perfect palindrome. The UAS2 of *CYC1* does not contain a repressor-binding site however it is possible that the inverted repeat within the *LPD* gene is in some way related to the repressor binding site. If so, this would represent an example of two overlapping regulatory elements. The role of such overlapping sites in the regulation of gene expression is discussed below.

The sequence AACTTA(A/G)TG present twice at positions -177 and -166 shows some homology to a 22 bp perfect dyad sequence, TCTCCAACTTATAAGTTGGAGA, present in the upstream region of the *ADHII* gene. This 22 bp sequence represents the binding site of the positive activator ADR1 involved in the induction of the *ADHII* gene during growth on non-fermentable carbon sources (Shuster *et al.*, 1986). The sequences at -177 and -166 are interesting as they represent a direct repeat in comparison to the inverted repeat in the *ADHII* promoter. This, therefore, represents another example of homology between a sequence in the upstream region of the *LPD* gene and a *cis*-acting element involved in mediating the catabolite repressible nature of another gene.

Finally, the sequence CACGTGA, present twice within a 23 bp inverted repeat, at positions -279 and -268 is identical to that shown recently to be required for mRNA initiation at the *TRP1* promoter (Mellor *et al.*, in press). The same motif has been identified near several other eukaryotic transcription units, including *GAL2* and the adenovirus major late promoter (Bram *et al.*, 1986; Sawadogo and Roeder, 1985). A very similar sequence,

TCAC(G/A)TG, is the core of a consensus sequence, found within all yeast centromeres, and represents the binding site of a protein termed centromere-binding protein or CP1. Interestingly the 23 bp repeat containing the two CACGTGA sequences is represented in a symmetrically truncated form at -361. The sequence TCACATGA at position +75 also shows homology to both the *TRP1* promoter sequence and the CP1 binding site (Bram *et al.*,, 1987). These regions of homology are shown in figure 3.8. A possible role for these sequences in the regulation of the *LPD* gene is discussed in chapter 5.

The significance of these regions of homology has been investigated and the results are presented in Chapters 4 and 5. The results show that several proteins appear to be binding to the upstream region of the *LPD* gene. The regions of interest discussed above are in several cases, in very close proximity to each other and some actuallyoverlap. This situation has been found in the promoters of other yeast genes. Both UAS sequences in *CYC1* bind several regulatory proteins (Pfeifer *et al.*, 1987; Forsburg and Guarente, 1988), and the *GAL1-GAL10* divergent promoter has been described as 'a battleground for positive and negative regulatory factors' (Finley and West, 1988). The presence of very close or overlapping binding sites for different regulatory factors may represent a common feature of eukaryotic promoters. The competitive binding of regulatory factors to adjacent or overlapping sites could be envisaged as a mechanism by which the effects of different control networks are integrated to correctly modulate gene expression.

CHAPTER 4 REGULATION OF THE *LPD* GENE BY GENERAL AMINO ACID CONTROL

4.1 GENERAL AMINO ACID CONTROL

The upstream region of the *LPD* gene contains three sites showing strong homology to the sequence identified as the GCN4 binding site. GCN4 is the most direct *trans*-acting positive regulator in the control network known as general amino acid control (Hope and Struhl, 1985). A consensus sequence for the GCN4-binding site has been established by comparing the regions upstream of genes known to be subject to general amino acid control (Hill *et al.*, 1986). This is shown in figure 4.1. Hinnebusch *et al.* (1985) have shown a 16 bp double stranded oligonuceotide with the sequence CTGACTCACGTTTTTG is sufficient to place the *CYC1* gene under some degree of general amino acid control. Furthermore, two copies of this oligonucleotide give derepression levels nearly as high as those observed when the entire 5' non-coding region of the *HIS4* gene is placed upstream of the *CYC1* gene. Thus the flanking sequence around the GCN4 binding site does not seem to be critical in determining the activity of this *cis*-acting element. These results suggest the presence of one or more sites, showing strong homology to the GCN4 binding site, in the upstream region of a yeast gene represents good evidence that the gene will be subject to general amino acid control.

Until recently, all those genes shown to be subject to general amino acid control were involved directly in the synthesis of amino acids or charged tRNA species. Regulation of lipoamide

Gene	Sequence
HIS1	AGCGTGACTCTTCCCGGAA
HIS1	GAGGTGACTCACTTGGAAG
HIS3	CGGATGACTCTTTTTTTTT
HIS4	ACAGTGACTCACGTTTTTT
ARG3	GTCGTGACTCATATGCTTT
ARG4	TGAATGACTCACTTTTTGG
CPA1	TTCTTGACTCGTCTTTTCT
CPA2	CGAATGACTCTTATTGATG
TRP5	AGAATGACTAATTTTACTA
TRP3	TCGTTGACTCATTCTAATC
TRP2	TTGCTGACTCATTACGATT
ILV1	GAGATGACTCTTTTTCTTT
ILV2	GCGATGATTCATTTCTCTG
LEU1	TAGATGACTCAGTTTAGTC
LEU4	TAAGTGACTCAGTTCTTTC
ILS1	ATGATGACTCTTAAGCATG
consensus	--rrTGACTCattt---t-
LPD1	CACGTGACTCGTTTTTAGA
LPD1	ATGATGAATCGTTTTTAAT
LPD1	CTTTTGACTCACCTCGAAT

Figure 4. 1. Consensus for the general control regulatory site and sequences from the *LPD* gene showing homology to this consensus.

DNA sequences for 16 prospective sites from 15 genes under general control. The TGACTC sequences are aligned and shown in bold print. The consensus line shows highly conserved nucleotides as upper case letters, conserved residues as lower case letters and nonconserved residues as dashes. The three sequences from the *LPD* gene showing homology to the GCN4 binding site are shown below the consensus sequence. Diagram modified from Hill *et al.*, (1986).

dehydrogenase by general amino acid control would thus represent an exception within this regulatory network.

General amino acid control, first identified by Delforge *et al.* (1975), refers to the derepression of a group of genes encoding more than 30 enzymes, involved in 9 different amino acid biosynthetic pathways (reviewed by Hinnebusch, 1988). The response is induced by the starvation of yeast cells for any one of at least 10 different amino acids. It remains possible that starvation for any amino acid is sufficient to elicit the response.

Genetic analysis has led to the identification of two classes of genes which are directly involved in general amino acid control. Firstly, *GCN* genes whose products are required for the derepression response and secondly, *GCD* genes whose products are required to maintain repression in non-starvation conditions (Greenberg *et al.*, 1986). Detailed genetic analysis, of mutations in the various *GCN* and *GCD* genes isolated, has allowed the hierachy of the various gene products to be established (Hinnebusch, 1988). This is shown in figure 4.2. It is not yet clear whether all of these genes encode proteins directly involved in general amino acid control but it is likely that some have a more indirect role in the regulation of gene expression. For example, *GCD1*, *GCD7* and *GCD10-13* have been shown to regulate expression of amino acid biosynthetic genes by controlling the steady-state level of their mRNA product (Harashima and Hinnebusch, 1986; Hill and Struhl, 1986; Myers *et al.*, 1986), while *GCD8* has been proposed to regulate structural gene expression through a post-transcriptional mechanism (Greenberg *et al.*, 1986).

As discussed in Chapter 1, GCN4 has been shown to encode a protein containing a DNA binding activity. This binding domain

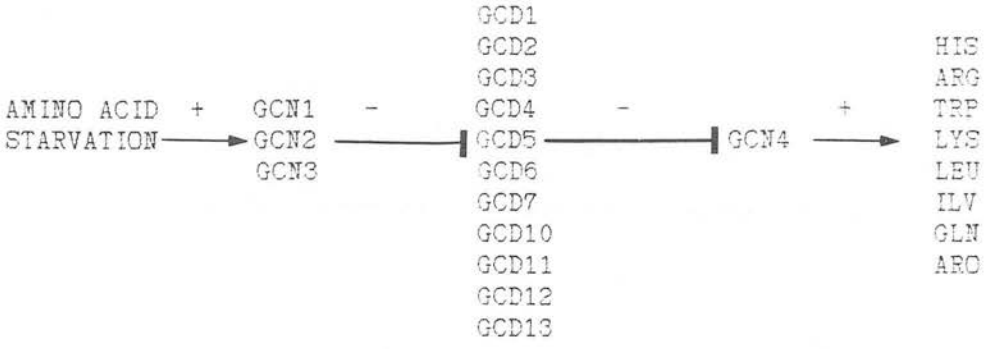


Figure 4. 2. Hierarchy of regulatory factors in general amino acid control.

Starvation for amino acid(s) results in the induction of the GCN1, GCN2 and GCN3 gene products which inhibit the activity of the GCD regulatory factors. The GCD gene products act to repress GCN4 synthesis which, when released from this repression, induces a wide range of amino acid biosynthetic genes (and possibly many others). Diagram from Hinnebusch, (1988).

interacts specifically with TGACTC regulatory sequences (Hope and Struhl, 1985). The protein also contains a short acidic region which is sufficient to activate transcription.

The regulatory mechanism, involved in the control of GCN4 activity, has not yet been completely determined. It is clear, however, that regulation of the protein involves the unusual nature of the mRNA species encoded by *GCN4* (Hinnebusch, 1984). The *GCN4* transcript is approximately 1500 nucleotides in length. A protein coding sequence of 281 codons is located near the 3' end of this transcriptional unit leaving approximately 600 nucleotides in the 5' leader. This leader contains four AUG codons followed, in each case, by either one or two sense codons and then a termination codon. Extensive point mutation studies have revealed that ORF's 3 and 4 have a negative regulatory effect on GCN4 synthesis and that these two ORF's are themselves negatively regulated by ORF 1 and 2 (Mueller and Hinnebusch, 1986). The negative effect of ORF 1 and 2 upon ORF 3 and 4 is in turn suppressed by the activity of the GCD1 gene product. A model summarizing the functional interactions between these upstream ORFs is shown in figure 4.3.

4.2 INVESTIGATION OF *LPD* TRANSCRIPT LEVELS DURING AMINO ACID STARVATION

In order to determine whether the *LPD* gene was regulated by general amino acid control *LPD* transcript levels were monitored during growth on conditions which induced this response. Several different methods have been reported for the induction of the general amino acid response. A strain, containing a leaky mutation

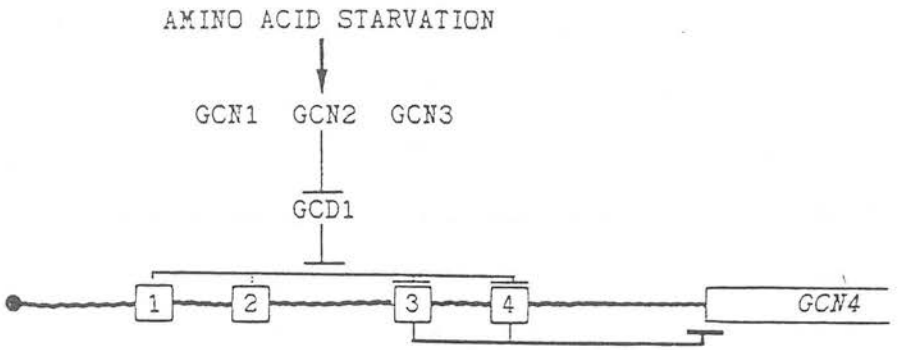
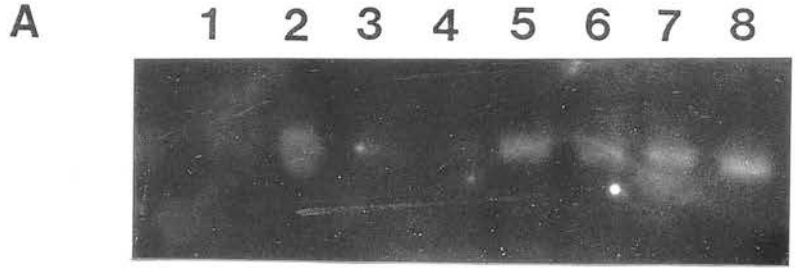


Figure 4. 3. *GCN4* mRNA open reading frame (ORF) interaction

Model summarizing functional interactions between the upstream ORFs in the *GCN4* mRNA leader that regulate translation of the GCN4 protein-coding sequences. ORFs 3 and 4 are required to efficiently reduce translation of the protein-coding sequences in repressing conditions, ORF1, and to a lesser extent ORF2, suppresses the blockade to initiation downstream exerted by ORFs 3 and 4. ORFs 1 and 2 are unable to exert this effect in repressing conditions when *GCD1* activity is present. Diagram from Hinnebusch, (1988).

in a structural gene for an amino acid biosynthetic enzyme, can be cultured on a medium lacking the required amino acid (Delforge *et al.*, 1975). Alternatively, wild-type cells grown initially on a media containing all 20 amino acids can be transfered to minimal medium. A lesser degree of derepression has also been observed in wild-type cells grown in the presence of amino acid supplements which lead to starvation for an amino acid other than those added as a result of regulatory interactions between related biosynthetic pathways (Niederberger *et al.*, 1981). Finally, amino acid starvation can be induced by the addition of an inhibitor of synthesis of one amino acid to wild-type cells growing on minimal media. Examples of such inhibitors include 5-methyltryptophan, which inhibits tryptophan biosynthesis by a false feedback effect upon anthranilate synthase (Schurch *et al.*, 1973) and 3-amino-1,2,4-triazole (3AT), which blocks histidine biosynthesis by competitively inhibiting imidazole glycerolphosphate dehydratase (Klopotowski and Wiater, 1965).

This last method was chosen to examine the effect the general amino acid response had upon *LPD* mRNA levels. PolyA$^+$-RNA samples were isolated from an exponentially growing culture of wild-type cells at time points before and after the addition of 3AT. Northern blot and slot blot results, showing the level of *LPD* mRNA at the time points taken, are shown in figure 4.4. The results show the level of *LPD* mRNA increase significantly, in both experiments, after the addition of the amino acid analogue. Figure 4.5 shows a densitometric scan of the autoradiogam shown in figure 4.4B and represents the quantification of the increase in *LPD* mRNA following the addition of 3AT. The increase of approximately 4 fold is



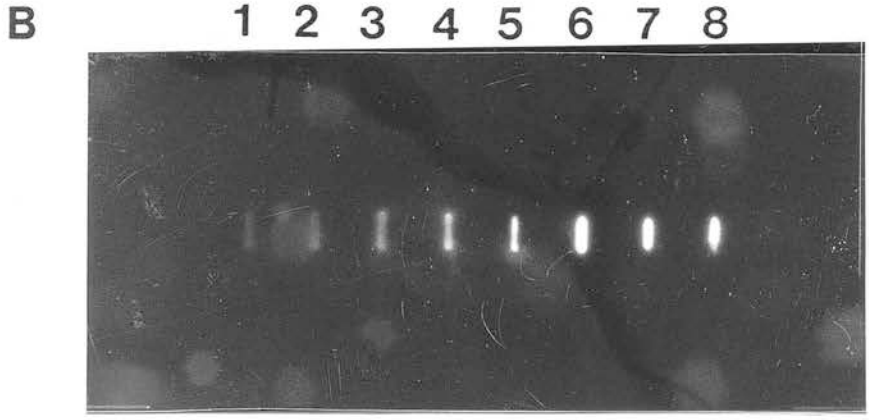


Figure 4. 4. *LPD* transcript analysis during amino acid starvation.

A. Northern showing *LPD* mRNA levels at 8 time points before and after the addition of 3AT at t = 0. Lane 1 = -120 min; 2 = -60 min; 3 = 0; 4 = 15 min ; 5 = 30 min; 6 = 60 min; 7 = 120 min; 8 = 180 min. 5 μg of polyA$^+$-RNA was loaded into each lane of a 0.8% agarose gel prior to electrophoresis.

B. Slot blot showing *LPD* mRNA levels before and after the addition of 3AT. Time points are the same as those in figure 4. 4A. 5 μg of polyA$^+$-RNA was loaded into each slot.

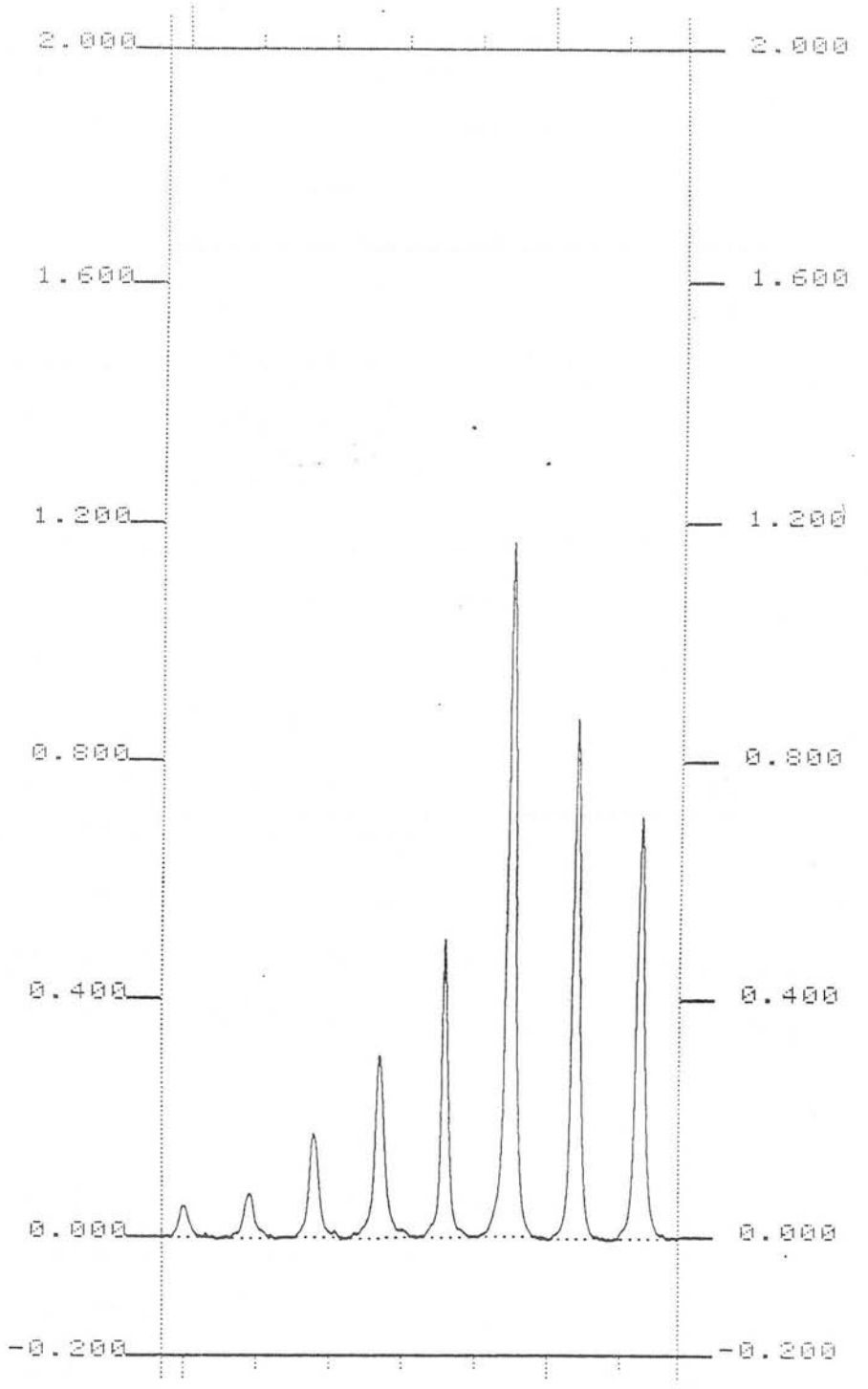


Figure 4. 5. Densitometric scan of the slot blot autoradiograph shown in figure 4. 5B.

typical of the extent of derepression observed for many other genes subject to general amino acid control, as is the peak in mRNA levels 1 hour after induction of the general amino acid response, (reviewed by Jones and Fink 1982; Penn *et al.*, 1984), and thus support the conclusion that the *LPD* gene is controlled by this regulatory network.

In the transcript analysis experiments, described here, the level of a non-derepressed transcript, such as *URA3*, has not been included as a loading control. Poly(A)$^+$-RNA levels, for each time point, were instead based simply upon spectrophotometric measurements. This is an important omission as the presence of contaminants, in the poly(A)$^+$ sample, can interfere with the quantification of RNA making quantitative comparisons between separate RNA samples unreliable. A more detailed analysis of the affect of amino acid starvation upon *LPD* transcript levels, using *URA3* as a negative control and *HIS3* as a positive control, is at present being carried out.

4.3 ANALYSIS OF THE BINDING OF GCN4 TO THE UPSTREAM REGION OF THE *LPD* GENE.

Regulation of lipoamide dehydrogenase by general amino acid control requires that GCN4 binds to the upstream region of the *LPD* gene. This was investigated with GCN4 protein which was synthesised *in vitro* and labelled with [^{35}S]-methionine. The protein was synthesised by transcribing the GCN4 gene using an SP6 expression vector containing the structural gene for GCN4, kindly supplied by K. Struhl and I. Hope. The transcript was then translated *in vitro* using a wheat germ extract to which [^{35}S]-

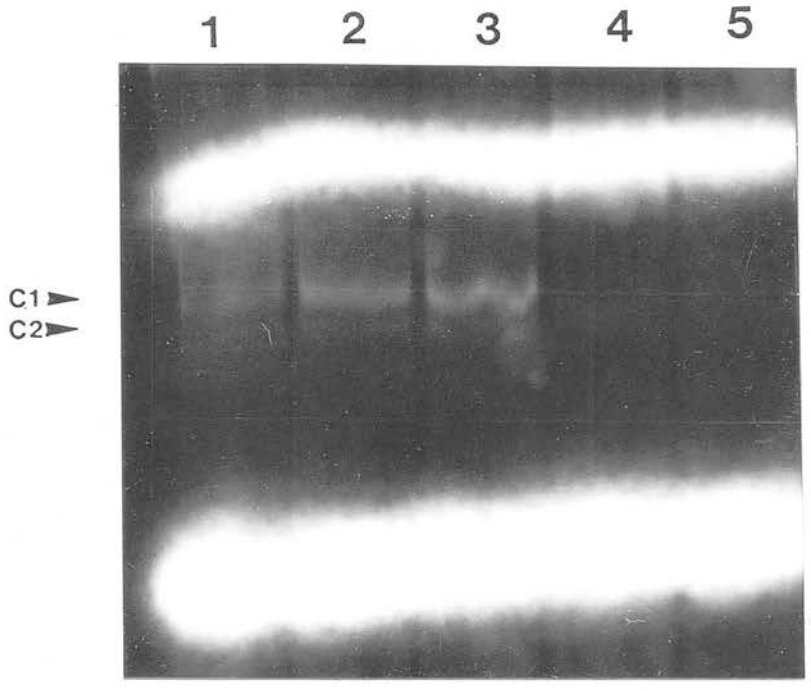


Figure 4. 6. Gel retardation analysis of the upstream region of the *LPD* gene using *in vitro* synthesised GCN4 protein.

[^{35}S]-methionine labelled GCN4 protein was synthesised *in vitro* as detailed in Chapter 2. The reaction extract, (2 µl), was added (without further purification) in each case to 18 µl DNA-binding buffer (Hope and Struhl, 1985), containing differing quantities of a 1020 bp *SalI/KpnI* fragment, covering the region from -831 to +190 within the *LPD* gene, digested with either *TaqI* (T) or *HinfI* (H). Lane 1 = 10 ng T; 2 = 20 ng T; 3 = 30 ng T; 4 = 30 ng H; 5 = no DNA. C1 and C2 represent protein/DNA complexes discussed in the text. The intense bands above and below the C1 and C2 bands are also discussed in the text.

methionine was added. *TaqI* and *HinfI* DNA fragments containing the upstream region of the *LPD* gene were incubated with labelled GCN4 and assayed for DNA/protein complex formation by gel retardation. The details of this experiment are shown in Fig 4.6

On increasing the amounts of *TaqI* DNA fragments in the incubation there was a corresponding increase in the intensity of two bands on the retardation gel designated C1 and C2 (lanes 1-3, Fig 4.6). The relative intensity and positions of C1 and C2 suggest that they might represent the 165 bp and 114 bp *TaqI* fragments, respectively, bound to GCN4 protein. The 165 bp fragment would be expected to bind the labelled protein more tightly, as it contains the two sites showing greatest homology to the GCN4 binding site. The *HinfI* fragments, in lane 4, show no detectable bands representing protein/DNA complexes. This is in agreement with the proposed location of the GCN4 binding sites as *HinfI* cuts at the sequence GANTC and will thus disrupt all three sites. A similar result was obtained when the upstream region of the *HIS3* gene, which contains 3 GCN4-binding sites, was digested with *HinfI* and assayed for GCN4-binding ability (Hope and Struhl, 1985) (see section 4.4). The intense bands, present in all five lanes, above and below the C1 and C2 bands represent unbound GCN4 protein and unincorporated [^{35}S]-methionine respectively; the altered migration of GCN4 when bound to DNA fragments is due to the altered charge conferred upon the complex by the bound DNA molecule (Ian Hope, personal communication).

In order to further investigate the binding of GCN4 to the upstream region of the *LPD* gene, DNase I footprinting analysis was carried out. A 172 bp *Mae*III/*Sst*II fragment, representing the

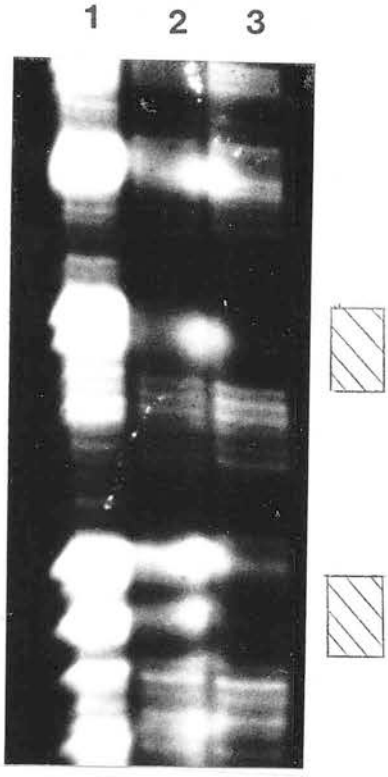


Figure 4. 7. Protection of *LPD* DNA from DNaseI cleavage by GCN4 protein.

Approximately 1 ng of a 172 bp *Mae*III/*Sst*II fragment (-354 to -183), end-labelled at the *Mae*III terminus, was incubated with various quantities of GCN4 protein, without further purification from the *in vitro* translation reaction. After controlled digestion with DNaseI, the products were examined on a sequencing gel alongside a dideoxy-sequencing ladder which provided size markers (not shown). Lane 1 = no GCN4; 2 = 2 µl translation products; 3 = 5 µl translation products. The two regions of supressed DNaseI cleavage are indicated alongside the cleavage products by hatched boxed and correspond with the two GCN4 binding sites at positions -264 and -247. The strand shown is the non-sense strand. There appears to be some DNase activity in the translation mix resulting in the increased degradation of the 172 bp fragment with increased amounts of translation products.

upstream region of the *LPD* gene from -357 to -185 and containing the two upstream GCN4 binding sites, was used in this analysis. The result of this experiment is shown in figure 4.7 The two regions of protection identified in the figure correspond to the location of the two regions showing homology to the GCN4 binding site within this 172 bp fragment.

4.4 COMPARISON OF GEL RETARDATION AND DNase I FOOTPRINTING RESULTS WITH PUBLISHED WORK.

Hope and Struhl were the first to report the synthesis of GCN4 protein *in vitro* and to demonstrate the binding of this protein to the upstream region of a gene subject to general amino acid control Hope and Struhl, (1985). GCN4 protein was generated by *in vitro* transcription and translation of the vector pSP64 containing the *GCN4* gene (pSP64-GCN4). Gel retardation analysis was then carried out using radio-labelled GCN4 and different restriction digests of a pUC8 derivative containing the *HIS3* gene (pUC8-*HIS3*) which were rendered blunt-ended using the large fragment of *E. coli* DNA polymerase I. In a typical reaction 0.5 µl of a 30 µl translation reaction was incubated with a 9 nM concentration of DNA fragments prior to gel retardation analysis.

By this method of analysis it was demonstrated that the radio-labelled protein was binding specifically to an 80 bp region in the upstream region of the *HIS3* gene containing one perfect and one imperfect TGACTC sequence. Also *HinfI* cut pUC8-*HIS3* showed no binding to GCN4. This enzyme cuts at GANTC sequences and would therefore cut at the sequence TGACTC.

An important series of controls in the gel retardation experiments, performed by Hope and Struhl, involved incubating the ^{35}S-translation products generated by transcription and translation of pSP64-*GCN4* and pSP64 DNA, and in the absence of exogenous RNA, with pUC9 DNA and with no DNA as well as with pUC8-*HIS3*. A single band, representing the complex formed between GCN4 protein and *HIS3* DNA, was detected only when the ^{35}S-translation products generated by transcription and translation of pSP64-GCN4 were incubated with pUC8-*HIS3*.

The location of the GCN4 binding site was further investigated using a series of pUC8-*HIS3* derivatives in which regions of the *HIS3* upstream region had been deleted. These plasmids were assayed for GCN4 binding ability as above and a 20 bp region was identified as the binding site of the protein. This 20 bp region contained a perfect copy of the TGACTC sequence. It was also noted that plasmids which lacked the perfect TGACTC sequence but retained an imperfect copy appeared to bind GCN4 protein at a level above that for nonspecific interactions suggesting this second site might represent a weak GCN4 binding site. The upstream regions of 8 other yeast genes were also tested for their ability to bind to GCN4 using the above assay. DNA fragments from the upstream regions of 4 of the genes, known to be subject to general amino acid control, bound to GCN4 while the fragments isolated from the other 4 genes, all of which were known not to be subject to general amino acid control, failed to bind to the protein.

In order to determine exactly where GCN4 protein was binding to the *HIS3* upstream region DNase I footprinting was carried out using the *in vitro* synthesised protein. In these experiments 1 ng

of assymetrically labelled DNA fragment was incubated with different amounts of GCN4 translation products as in the gel retardation experiments discussed above. The protein/DNA complexes were then subjected to a controlled digestion with DNase I and the resulting DNA fragments analysed on an acrylamide sequencing gel. Protection experiments using two *HIS3* upstream fragments demonstrated that GCN4 protein protected a 10 bp region on one strand and a 5 bp region on the other strand and that this region included the TGACTC sequence within the 20 bp region determined by gel retardation analysis.

Studies involving saturation mutagenesis of the yeast *HIS3* regulatory site followed by gel retardation analysis using *in vitro* synthesised GCN4 (Hill *et al.*, 1986) have provided a clearer picture of the optimal sequence required for GCN4 binding. The palindromic sequence ATGACTCAT was shown to represent the optimal binding site for GCN4 which is in good agreement with the consensus sequence RRTGACTCATTT, determined by comparing several GCN4 binding sites. A similar conclusion was reached using GCN4 protein obtained from an overproducing strain of *E. coli* (Arndt and Fink, 1986). Using higher concentrations of GCN4 protein than were available from *in vitro* synthesis they showed the protein would protect all 'TGACTC-like' sequences in the upstream regions of the *HIS3*, *HIS4*, *ILV1* and *ILV2* genes during DNase I footprinting experiments. Analysis of the relative binding constants of the different sequences which GCN4 protected led to the conclusion that the sequence RRTGACTC, followed by a short stretch of thymidine residues, represented the optimal binding site for GCN4. Arndt and Fink did not conclude that the binding of GCN4, when present in higher concentrations, was

artifactual. Previous genetic studies had shown three of the TGACTC sequences from the upstream region of *HIS4*, identified as binding to GCN4 only when the protein was present in high concentrations, were important for general control. These results suggest relatively high concentrations of GCN4 protein are needed to identify all those sites to which the protein binds in the upstream region of any gene subject to general amino acid control.

The experiments using *in vitro* synthesised GCN4 presented in this thesis were based upon those described by Hope and Struhl discussed above. GCN4 protein was both synthesised and assayed for its ability to bind to the upstream region of the *LPD* gene using materials and conditions similar to those previously reported. A number of the controls used in the experiments discussed above were however omitted from those reported here and it is important to recognise the absence of these controls before attempting to draw conclusions from the results obtained.

1. Incubation of DNA fragments with translation products in which pSP64 replaced pSP64-*GCN4* in the initial *in vitro* transcription and in which no RNA was included for *in vitro* translation. These two controls would have clearly determined whether any product of the *in vitro* transcription/translation reaction other than GCN4 protein was responsible for the two bands seen in lanes 1 to 3 in figure 4.6 or for the regions of protection seen in figure 4.7.

2. In the gel retardation experiment individual *TaqI* fragments were not assayed for there ability to bind to GCN4. This would have clearly determined with which fragment the radio-labelled protein was binding.

3. DNA fragments containing the *HIS3* binding site were not included in the gel retardation assay as a positive control.

In addition to the above it should be noted that the concentration of DNA fragments used in the gel retardation experiment were clearly not in excess relative to the quantities of labelled GCN4 used thus differing from the conditions described by Hope and Struhl. Increased concentrations of DNA from lanes 1 to 3 resulted in an increase in the two bands observed. If these bands did correspond with protein/DNA complexes they would have remained at a constant intensity regardless of DNA concentration had the DNA been in excess relative to the labelled protein.

4.5 REGULATION OF LIPOAMIDE DEHYDROGENASE BY GENERAL AMINO ACID CONTROL

The three experiments described within this chapter were all designed to answer the question 'is lipoamide dehydrogenase subject to general amino acid control?'. Firstly, the relative levels of *LPD* mRNA were analysed before and after yeast cells were placed under conditions known to induce the general amino acid response. Secondly, the ability of *in vitro* synthesised GCN4 protein to bind to the upstream region of the *LPD* gene was assayed both by gel

retardation and by DNase I protection experiments. None of the results obtained represent conclusive proof to answer the above question. As discussed in sections 4.3 and 4.4 all three experiments lacked one or more important controls needed to provide an unambiguous result. Hence no firm conclusion can be drawn at this stage concerning the regulation of lipoamide dehydrogenase by general amino acid control. All three experiments presented have produced what can perhaps best be described as suggestive results which should, when taken together, encourage further investigation to clearly determine whether the *LPD* gene is subject to general amino acid control.

The upstream sequence of the *LPD* gene itself represents good circumstantial evidence that lipoamide dehydrogenase is subject to general amino acid control. Previous reported work, reviewed in this chapter, suggests the presence of several 'TGACTC-like' sequences in the upstream region of a gene, as is the case with the *LPD* gene, should be sufficient to place that gene under general amino acid control. One experiment in particular, mentioned at the start of this chapter, in which it was shown TGACTC-containing sequences were sufficient to place a gene, not normally subject to this response, under general amino acid control imply the presence of the sequences identified upstream of the *LPD* gene should be sufficient to place it under general amino acid control.

In conclusion it can only be said, from the results presented within this chapter and from the sequence data presented in chapter 3, that the *LPD* gene and thus lipoamide dehydrogenase in *Saccharomyces cerevisiae* represents 'a good candidate' for an enzyme

subject to the general amino acid response and one worthy of further study to resolve this issue.

4.6 THE ROLE OF GENERAL AMINO ACID CONTROL IN YEAST

Regulation of lipoamide dehydrogenase by general amino acid control would perhaps not be totally unexpected. As discussed in Chapter 1 the enzyme plays an important role in the TCA cycle. 2-Oxoglutarate and oxaloacetate, both TCA cycle intermediates, are the substrates for the glutamate and aspartate families of amino acids, respectively. Thus, indirectly, the genes which encode enzymes within the TCA cycle are involved in amino acid biosynthesis. The co-ordinated derepression of these TCA cycle genes with those directly involved in amino acid biosynthesis can therefore be envisaged.

Unfortunately, at present, the data bases contain no upstream sequence data for the other yeast genes of the TCA cycle. There is one report that NAD-linked glutamate dehydrogenase and fumarase are derepressed in arginine bradytrophs (Delforge *et al.*, 1975); this, however, represents the only other report suggesting that a TCA cycle enzyme may be subject to general amino acid control. It therefore remains to be seen whether other TCA cycle genes from *S. cerevisiae* contain GCN4 binding sites and are subject to general amino acid control.

There have recently been three separate reports identifying the sequence TGACTC in the upstream region of the genes involved in purine biosynthesis (Gedvilaite *et al.*, 1988; Lahti *et al.*, 1988; Mvasnikov and Smirnov, 1988). The genes *ADE1*, *ADE2*, *ADE3*, *ADE4*, *ADE5* and *ADE7* all contain this six base pair sequence in their

5'-noncoding regions. Data investigating the functional significance of these sequences are not yet available but the presence of GCN4 binding sequences in all of these genes strongly indicates purine biosynthesis is also regulated by general amino acid control.

The general amino acid response induces a 3- to 5-fold derepression of amino acid biosynthetic genes (Jones and Fink, 1982). This is in comparison with derepression of 20 to 180 fold for many of the amino acid specific control systems present in yeast. In the absence of a particular amino acid the pathway specific regulation would thus mask the derepression by general amino acid control, of genes within that pathway. It has also been shown recently that the general amino acid response can be induced by growth conditions which trigger the heat shock response (Messenguy and Scherens, 1988). Heat shock response, first identified in *Drosophila melanogaster* (Ashburner *et al.*, 1979), and now shown to operate in most, if not all, organisms (Schlesinger, 1982; Pelham, 1985), represents another regulatory network that is induced by a stress signal. Furthermore, *gcd* and *gcn* mutants affected in general amino acid control regulation are also affected in their thermotolerance (Messenguy and Scherens, 1988).

The low levels of derepression induced by the general amino acid response, the wider role this regulatory network is now being shown to have, and the range of conditions which appear to induce the response, all suggest general amino acid control has a more 'global' effect upon cellular physiology than that previously envisaged. The general amino acid control network may eventually prove to be part of a more extensive response to stress which includes heat shock response and perhaps other as yet undefined

control elements which modulate the expression of a wide range of genes. The inclusion of the gene encoding lipoamide dehydrogenase and indeed the other TCA cycle genes within such a control network does not seem unlikely given the important role the cycle plays in both the generation of ATP and the production of intermediates used in a variety of biosynthetic reactions.

CHAPTER 5 DNA-BINDING ACTIVITIES SPECIFIC FOR THE *LPD* GENE

In Chapter three the identification of a number of sites showing homology to known yeast regulatory motifs, within the 5' noncoding region of the *LPD* gene, was discussed. These observations prompted a search for *trans*-acting factors which bind to these sites and modulate the expression of the *LPD* gene.

The strategy used to search for DNA-binding activities was based upon one used by Ruet *et al.* (1984). DNA-binding proteins were isolated from a yeast crude extract, prepared from YEPG grown cells, using an heparin-Sepharose affinity column. The protein fractions, each containing a complex mixture of proteins, were then assayed for specific interactions with the *LPD* gene by gel retardation.

5.1 SEARCH FOR SPECIFIC DNA-BINDING ACTIVITIES

The initial gel retardation studies involved a 1022 bp *SalI/KpnI* fragment from -831 to +190. This was cut with *Sau3AI* to give 3 fragments 648, 196 and 178 bp in length which were end-labelled with [α-^{32}P]dCTP prior to incubation with the column fractions. A DNA-binding activity was detected which had peak activity in fraction 37. A retardation gel showing the mobility of *Sau3AI* fragments after incubation with increasing amounts of fraction 37 and fraction 31 is shown in figure 5.1. The retardation of the 648 bp fragment, by at least one DNA-binding activity in fraction 37, can clearly be seen. No retardation was observed with fraction 31. In this gel the two new bands (and several fainter bands detected on the original autoradiograph) observed on addition of fraction 37 to the labelled DNA, represent protein/DNA retardation complexes.

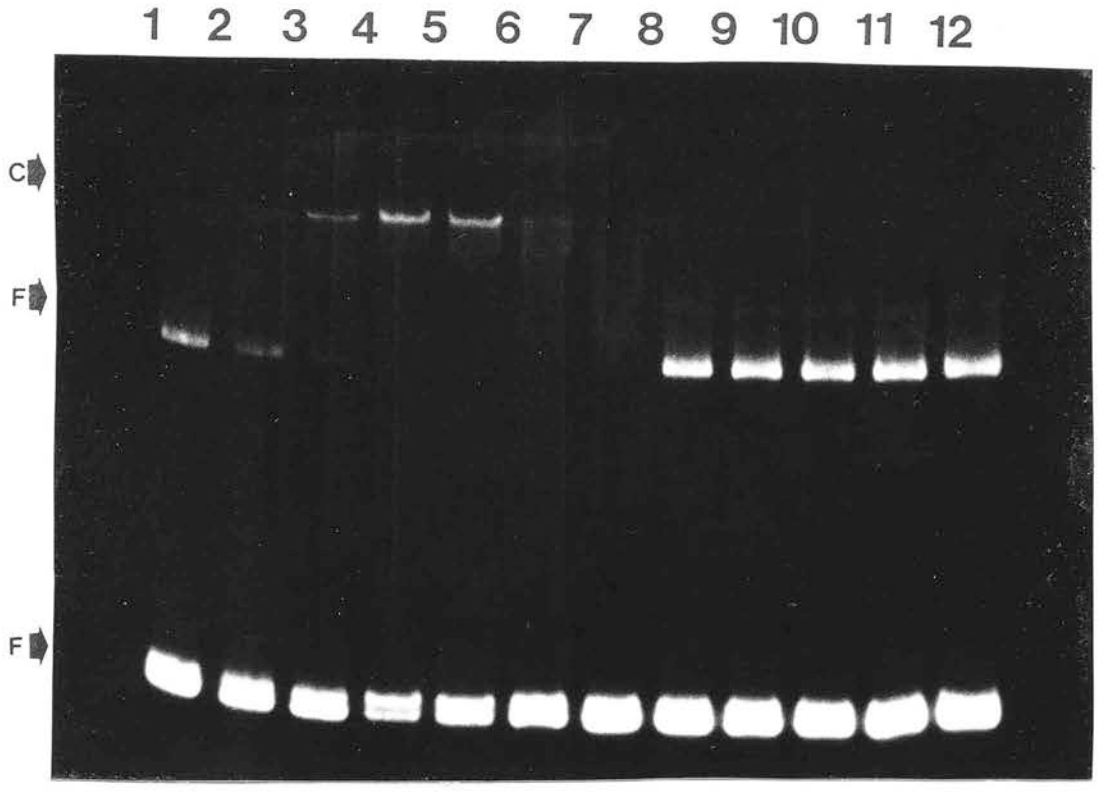


Figure 5. 1. Gel retardation analysis of proteins binding to the upstream region of the *LPD* gene.

The 1022 bp *SalI/KpnI* fragment (-831 to +190) was digested with *Sau*3AI to generate 3 fragments of 648 bp (-635 to +12), 196 bp (-831 to -636) and 178 bp (+13 to +190) which were end-labelled using [^{32}P]-dCTP. The mixture of fragments (containing 1 ng radio-labelled DNA) was incubated with differing amounts of fractions obtained by heparin-Sepharose chromatography and run on polyacrylamide gels according to Huet *et al.*, (1985). The cold carrier DNA was pBR322 with a *Deinococcus radiodurans* DNA insert. F, free DNA; C, protein/DNA complexes.

Lane	:	1	2	3	4	5	6	7	8	9	10	11	12
Fraction No.	:	-	37	37	37	37	37	37	31	31	31	31	31
Protein added (μg):		-	1	2	3	4	5	6	1	2	3	4	5

A separate experiment, in which the formation of retardation complexes was examined after the addition of increasing amounts of proteinase K to the fractions, confirmed the DNA-binding factors present in the fractions were proteins (data not shown).

5.2 FURTHER LOCALIZATION OF THE DNA BINDING SITES

In an attempt to localise the binding site of the protein retarding the largest of the three fragments assayed, the 648 bp *SalI/Sau3AI* fragment was digested with *Mae*III producing two fragments 280 bp (-635 to -355) and 367 bp (-354 to +12) in length. The result is shown in figure 5.2. The larger of the two fragments was clearly retarded by a protein in fraction 37 while the smaller fragment was not shifted suggesting the protein binding to the 648 bp fragment was binding downstream of the *Mae*III site.

The result of a gel retardation experiment, using three *HinfI* fragments covering the region from -245 to +190, is shown in figure 5.3. The two larger *HinfI* fragments, representing the regions +8 to +190 and -245 to -113, were retarded, to differing degrees, by increasing amounts of fraction 37. The smallest *HinfI* fragment representing a region from -112 to +7 appeared to be unaffected. The 1022 bp *SalI/KpnI* fragment was digested with *TaqI* producing 5 fragments which could be assayed individually for any specific protein interacting with each one.

The result of a gel retardation experiment (using a new set of protein fractions) involving all 5 *TaqI* fragments is shown in figure 5.4. This shows that a protein, which had peak activity in fraction 7, bound to both the 165 and 236 bp fragments representing the -383 to -219 and -45 to +190 regions respectively. In addition, the 447 bp

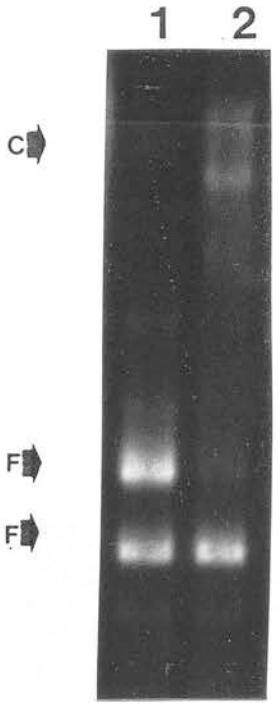


Figure 5. 2. Gel retardation using the 2 *Mae*III/*Sau*3AI fragments.

The 648 bp *Sau3AI* fragment was digested with *Mae*III to generate 2 fragments of 280 bp (-635 to -355) and 367 bp (-354 to +8). Conditions were the same as those in the legend to figure 5.1. F, free DNA; C, protein/DNA complexes.

Lane	1	2
Fraction No.	-	37
Protein added (μg):	-	5

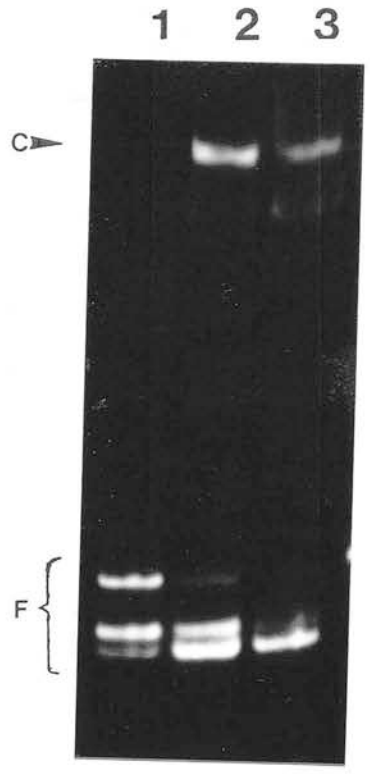


Figure 5. 3. Gel retardation using the 3 *Hinf*I fragments.

The 2 *Hinf*I fragments 133 bp (-245 to -113) and 120 bp (-112 to +7) in length and the 183 bp *HinfI/KpnI* (+8 to +190) were incubated with increasing amounts of fraction 37 isolated from an heparin-Sepharose column. Conditions were the same as those in the legend to figure 5.1. F, free DNA; C, protein/DNA complexes.

Lane	:	1	2	3
Fraction No.	:	-	37	37
Protein added (μg):		-	1	2

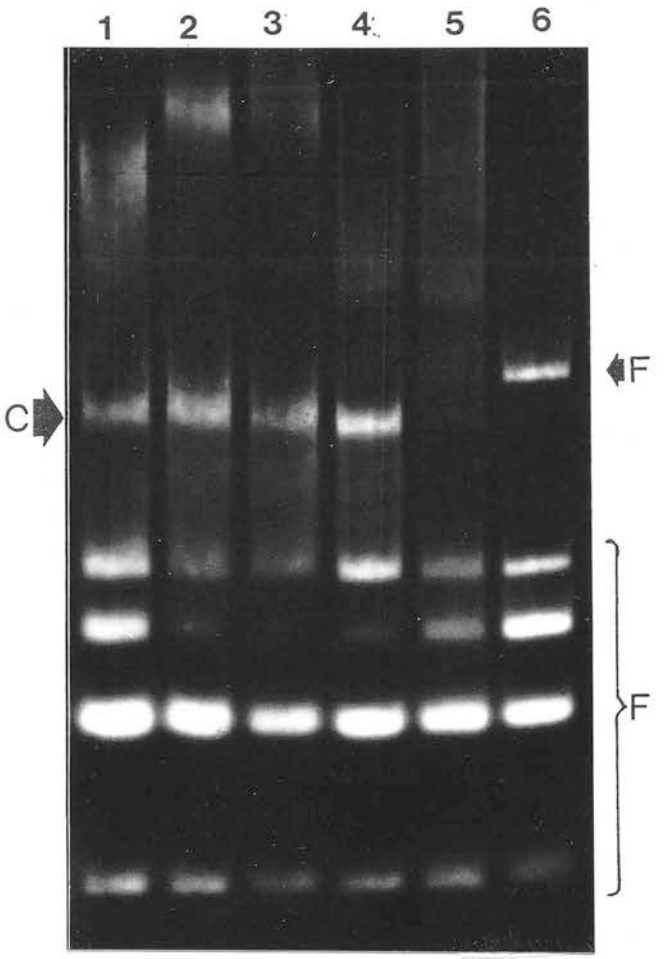


Figure 5. 4. Gel retardation analysis of proteins binding to the upstream region of the *LPD* gene.

The 1022 bp *SalI/KpnI* fragment (-831 to +191) was digested with *TaqI* producing five fragments 447 bp (-830 to -384), 236 bp (+190 to -45), 165 bp (-383 to -219), 114 bp (-218 to -105) and 59 bp (-105 to -46) in length which were end-labelled using [^{32}P]-dCTP. The mixture of fragments (containing 1 ng radio-labelled DNA) was incubated with different protein fractions. Conditions were the same as those in the legend to figure 5.1. F, free DNA; C, protein/DNA complexes.

Lane	:	1	2	3	4	5	6
Fraction No.	:	5	6	7	8	10	-
Protein added (μg):		2	2	2	2	2	-

fragment from -680 to -384 appears to have been retarded by all five of the protein fractions with which it was incubated.

5.3 USE OF INDIVIDUAL *TAQI* FRAGMENTS TO IDENTIFY THE SEPARATE DNA-BINDING ACTIVITIES.

In an attempt to localise the binding sites of the protein(s) identified in the above experiments a further set of fractions were isolated in a similar fashion to that described above. In this experiment, however, both the gradient of salt concentration and the fraction size were reduced in an attempt to improve the separation of distinct DNA-binding activities which may have been eluted together in the original fractionations. The absorbance profile of this heparin-Sepharose fractionation run are shown in figure 5.5. The following work was done in conjunction with Wendy Armstrong (B.Sc. honours student).

5.3.1 Results using the 114, 59 and 42 base pair fragments

The *Taq*I fragments discussed above (excluding the 447 bp fragment) along with a 42 bp *Taq*I fragment, representing the extreme 5' terminus of the sequenced region upstream of the *LPD* gene, were examined individually for their ability to interact with the DNA-binding activities isolated. The gel retardation experiments using the 114, 59 and 42 bp fragments were shown in figures 5.6, 5.7 and 5.8 respectively. All three fragments are retarded to varying degrees by a protein, which peaks at fraction 34. The retardation observed with the 114 bp fragment, however, was much more extensive than that seen with the other two fragments. This fragment contains the sequence showing homology to the HAP2 and HAP3 binding site identified by Forsburg and Guarente, (1988). The DNA-binding activity interacting with the 114 bp fragment could therefore be either of these proteins. This idea is supported by the method used to

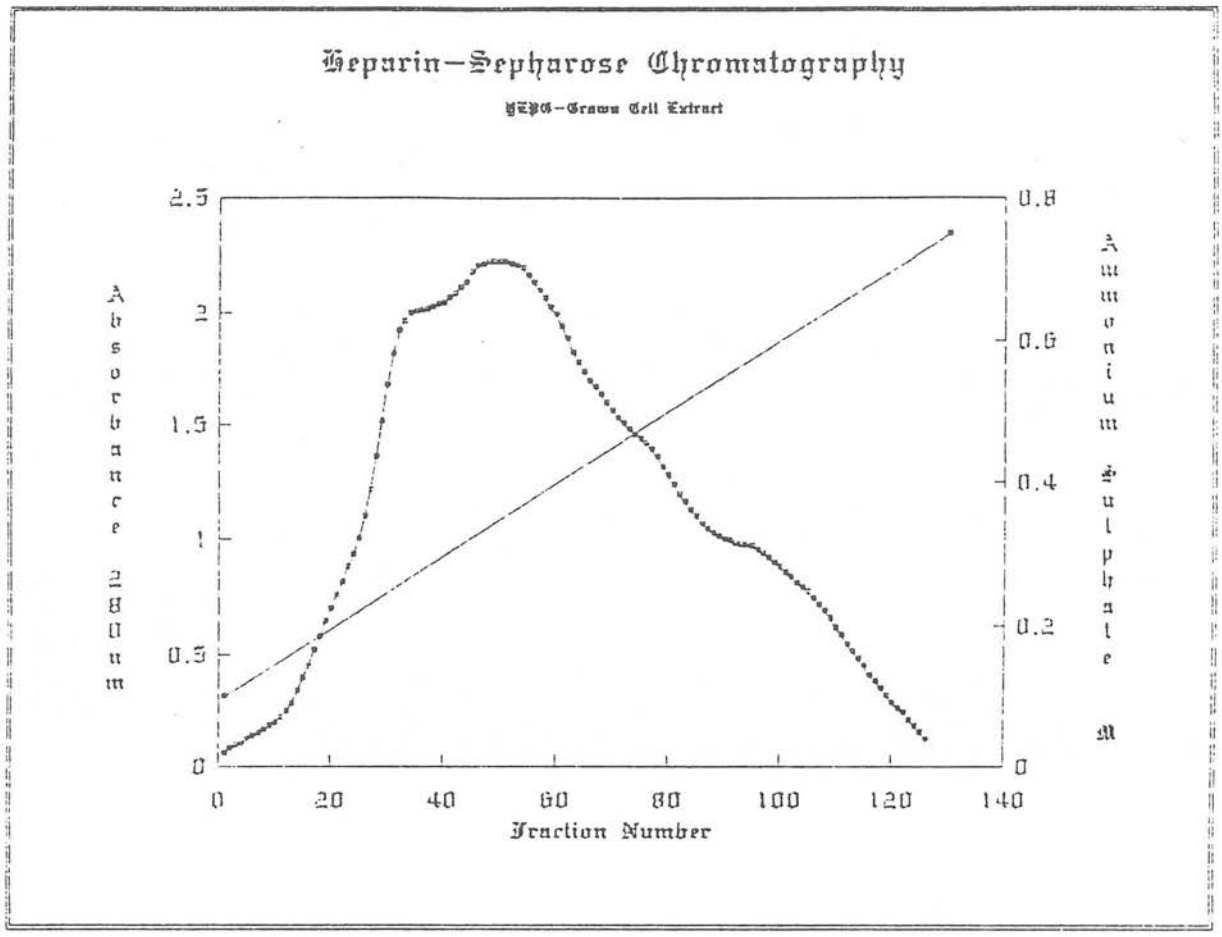


Figure 5. 5. Heparin-Sepharose chromatography of a crude yeast extract.

An extract of cells of *S. cerevisiae* strain DC5 grown on YEPG medium (20 l) was loaded onto a heparin-Sepharose column (100ml) as described in Chapter 2. The column was washed with column buffer containing 0.1 M $(NH_4)_2SO_4$ and then eluted with a linear gradient (0.1 to 0.75 M) of $(NH_4)_2SO_4$. 125 fractions (4 ml) were collected. These were pooled in pairs and stored for further analysis. The elution of material absorbing at 280 nm is shown for the gradient elution along with the salt gradient.

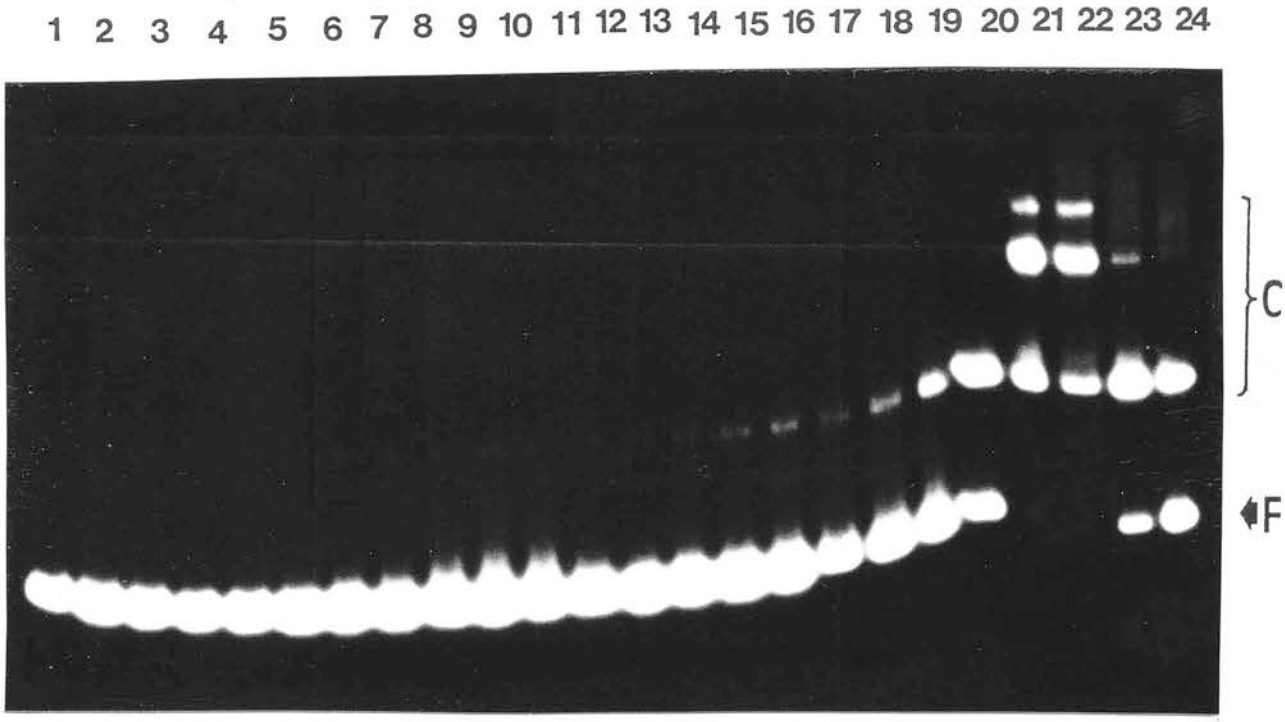


Figure 5. 6. Gel retardation of the 114 bp fragment.

The 114 bp *TaqI* fragment (-218 to -105) was incubated with different protein fractions isolated by heparin-Sepharose chromatography, as in figure 5.5. Each sample includes 0.5 ng radio-labelled DNA, 0.525 µg protein, 1 µg poly(dI): poly(dC) carrier DNA, incubated in 1X footprinting buffer (Huet *et al.*,1985), for 15 min at 25°C, in a total volume of 15 µl prior to electrophoresis on polyacrylamide gels according to Huet *et al.*, (1985).. F, free DNA; C, protein/DNA complexes.

Lane	:	1	2	3	4	5	6	7	8	9	10	11	12
Fraction:		-	2	4	6	8	10	12	14	16	17	18	19

Lane	:	13	14	15	16	17	18	19	20	21	22	23	24
Fraction:		19	20	21	22	23	24	26	28	30	32	34	36

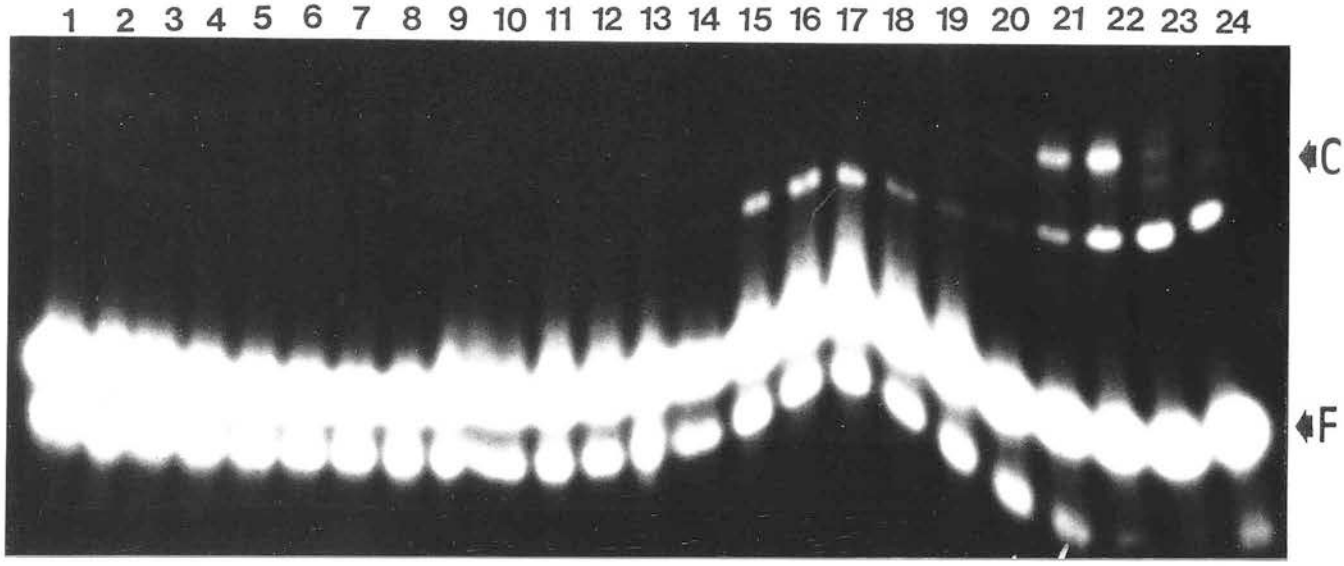


Figure 5. 7. Gel retardation of the 59 bp *Taq*I fragment.

The 59 bp *Taq*I fragment (-105 to -46) was incubated with different protein fractions isolated by heparin-Sepharose chromatography, as in figure 5.5. Sample details as in figure 5.6. F, free DNA; C, protein/DNA complexes.

Lane	:	1	2	3	4	5	6	7	8	9	10	11	12
Fraction:		-	2	4	6	8	10	12	14	16	17	18	19

Lane	:	13	14	15	16	17	18	19	20	21	22	23	24
Fraction:		19	20	21	22	23	24	26	28	30	32	34	36

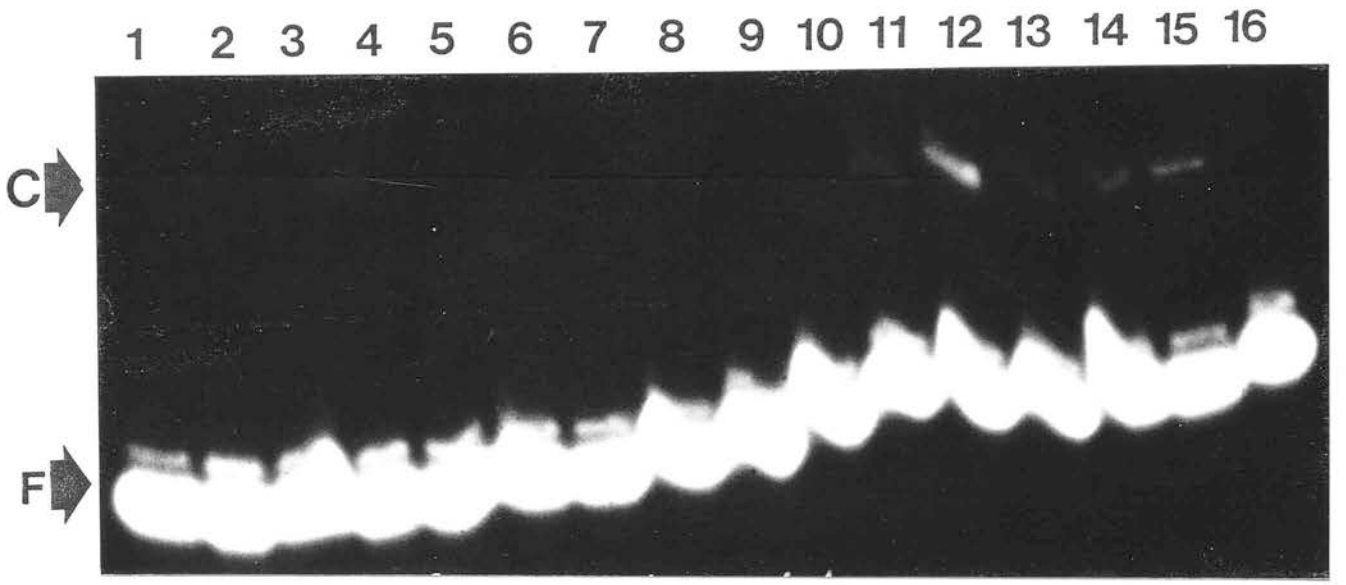


Figure 5. 8. Gel retardation of a 42 bp fragment.

A 42 bp *TaqI* fragment from plasmid pUC18 was incubated with different protein fractions isolated by heparin-Sepharose chromatography, as in figure 5.5. Sample details as in figure 5.6. F, free DNA; C, protein/DNA complexes.

Lane	:	1	2	3	4	5	6	7	8
Fraction	:	-	2	6	10	14	17	19	21

Lane	:	9	10	11	12	13	14	15	16
Fraction	:	23	26	30	34	38	42	46	50

prepare the fractions. Cells were grown on YEPG, a non-fermentable carbon source, which would induce the synthesis of these two proteins. In addition, it has been shown that HAP2 and HAP3 can be isolated from crude extracts of yeast cells grown on non-fermentable carbon sources (Forsburg and Guarente, 1988). The 114 bp fragment also contains the two sites showing some homology to the ADR1 binding site. By the same arguments as those presented for the HAP2 and HAP3 proteins, the ADR1 protein is a possible candidate for the protein binding to the 114 bp fragment as the regions of homology to the ADR1 binding site identified in the upstream region of the *LPD* gene are also present within this 114 bp fragment.

An alternative proposal to explain the results of figures 5.6, 5.7 and 5.8 is that the activity in fraction 34 represents a non-specific DNA binding protein, not totally competed off by the carrier DNA, which for some reason binds with a greater affinity to the 114 bp fragment. The DNA/protein complexes formed by the 114 bp fragment and, to a lesser extent, the 59 bp fragment are similar to the type obtained when a protein is binding to a fragment several times. An example of this is seen in the *E. coli* lac repressor-operator interactions described in one of the first papers to use gel retardation assays (Fried and Crothers, 1981). Increasing amounts of the lac repressor produced a 'ladder' of protein/DNA complexes as more protein bound to each DNA fragment.

The pattern of retardation complexes can be seen again in figure 5.9 in which the 114 bp and the 165 bp *TaqI* fragment were incubated with increasing amounts of fraction 34 and 19 respectively. The 165 bp fragment was retarded by fraction 19 (see below) to give two protein/DNA complexes and increasing amounts of the fraction did not alter this pattern. In contrast, the formation of several distinct complexes which

were retarded further with increasing amounts of fraction 34 was seen for the 114 bp fragment. This might indicate that the 114 bp fragment contains several sites for a regulatory protein as seen in the *GAL1-GAL10* promoter (St. John and Davis, 1984), or that the fragment is binding a non-specific DNA-binding protein more than once. A search for the presence of this DNA-binding protein in fractions isolated from a YEPD culture, which will contain much less HAP2 or HAP3 proteins, may resolve this issue.

5.3.2 Results obtained using the 165 and 236 base pair fragments

Figures 5.10 and 5.11 show the gel retardation results obtained using the 165 and 236 bp fragments. Both fragments were retarded by a DNA-binding activity with peak activity in fraction 19. Furthermore, both fragments show similar retardation complexes each consisting of two intense bands above two fainter bands. This is shown again in figure 5.12 in which both fragments were run on the same gel. In an experiment in which the amount of fraction 19 added to the 165 bp fragment was varied, shown in figure 5.9, no protein/DNA complexes were observed above the two main complexes identified with increasing amounts of the protein fraction. These results support the proposal that the same protein was binding to both fragments. Further evidence to support this conclusion came from an experiment in which the 165 and 236 bp fragments were mixed together prior to their incubation with fraction 19. This is shown in figure 5.13. The two fragments, individually bound to the protein in fraction 19 to give DNA/protein complexes of similar intensity. When the fragments were mixed the intensity of each band decreased indicating the two DNA fragments were competing for the same protein. The central band appeared stronger due to the additive effects of the upper and lower

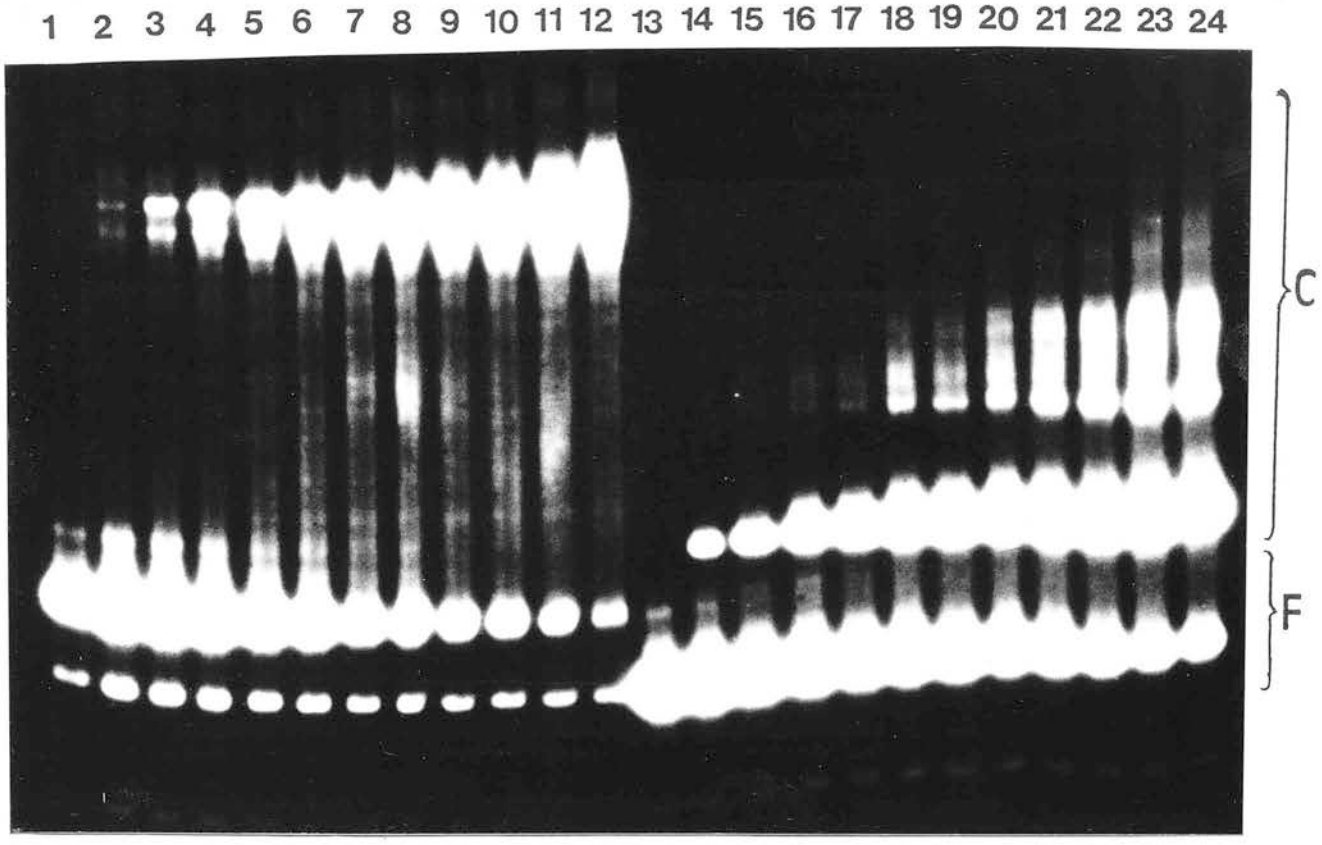


Figure 5. 9. Gel retardation of the 165 and 114 bp fragments.

The 165 bp (lanes 1 to 12) and 114 bp (lanes 13 to 24) *Taq*I fragments were incubated with increasing amounts of fraction 19 and fraction 34, respectively, isolated by heparin-Sepharose chromatography, as in figure 5.5. Sample details as in figure 5.6. F, free DNA; C, protein/DNA complexes.

Lane	:	1	2	3	4	5	6	7	8	9	10	11	12
Protein (μg):		0	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.8	1	1.5	2.0

Lane	:	13	14	15	16	17	18	19	20	21	22	23	24
Protein (μg):		0	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.8	1	1.5	2.0

bands of the 165 and 236 bp complexes respectively. The difference in fragment size was presumably responsible for the different mobility of the complexes, observed for the two fragments.

The 165 bp fragment contains two copies of the GCN4 binding site. GCN4, however, is unlikely to be responsible for the observed retardation complexes. The protein fractions were isolated from the crude extract of a yeast culture grown on YEPG. Under such growth conditions GCN4 synthesis is repressed (Hinnebusch and Fink, 1983). In addition, the concentration of GCN4 protein in the cells is known to be very low even during amino acid starvation conditions when the synthesis of the protein is induced (Ian Hope, personal communication). Antibodies, raised against GCN4 protein synthesised *in vitro*, have so far failed to detect its presence in yeast cells giving an indication of the low levels of GCN4 present (Ian Hope, personal communication). The retardation complexes are, therefore, almost certainly not due to GCN4 protein.

The 165 bp fragment also contains two copies of the sequence CACGTGA known to be involved in the regulation of other yeast genes (Mellor *et al.*, in press). As discussed in Chapter 3 this is very similar to the sequence TCAC(G/A)TG, identified as the core of a consensus sequence involved in yeast centromere structure and refered to as centromere DNA element I (CDEI) and representing the binding site of centromere-binding protein 1 (CP1), (Buchman *et al.*, 1988). Furthermore, the 236 bp fragment contains the sequence TCACATGA covering the region from +75 to +82 in the *LPD* gene. These sequences, therefore, represent a possible site at which the protein, present in fraction 19, was binding.

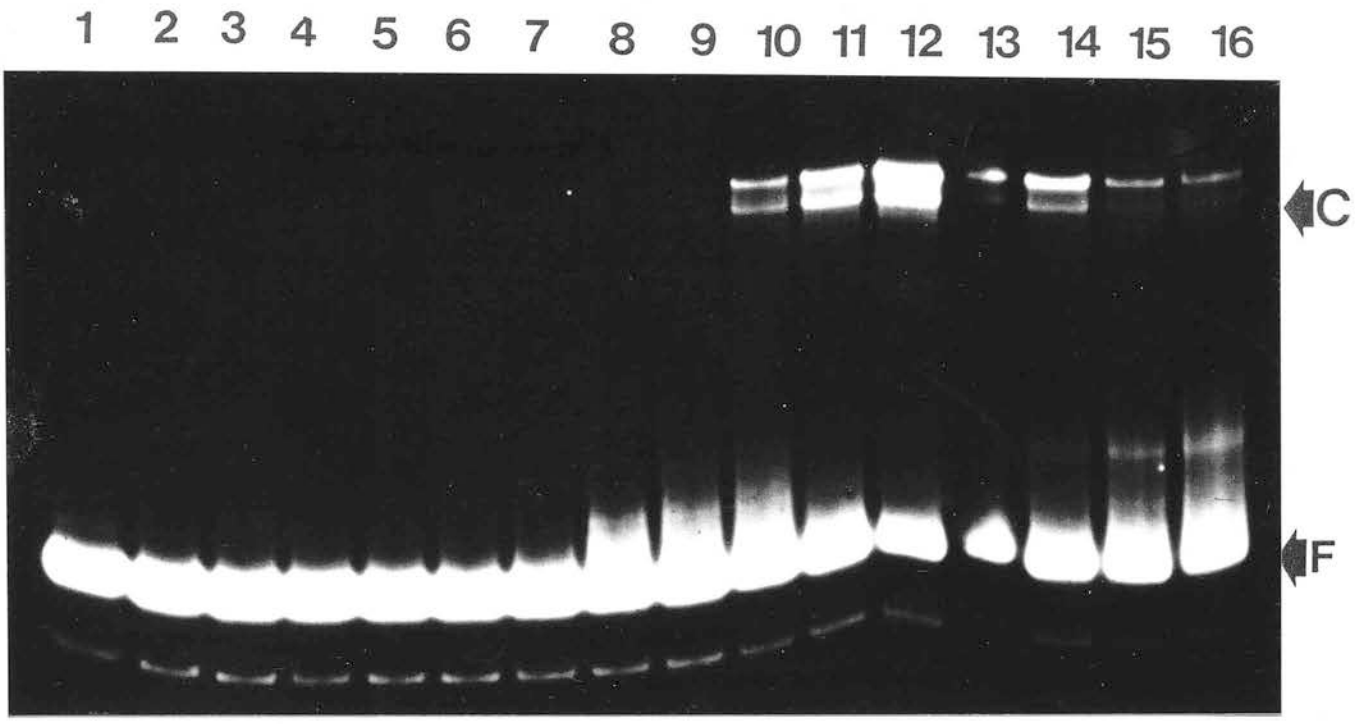


Figure 5. 10. Gel retardation of the 165 bp fragment.

The 165 bp *TaqI* fragment (-383 to -219) was incubated with different protein fractions isolated by heparin-Sepharose chromatography, as in figure 5.5. Sample details as in figure 5.6. F, free DNA; C, protein/DNA complexes.

Lane	:	1	2	3	4	5	6	7	8
Fraction	:	-	2	4	6	8	10	12	14

Lane	:	9	10	11	12	13	14	15	16
Fraction	:	16	17	18	19	20	21	22	23

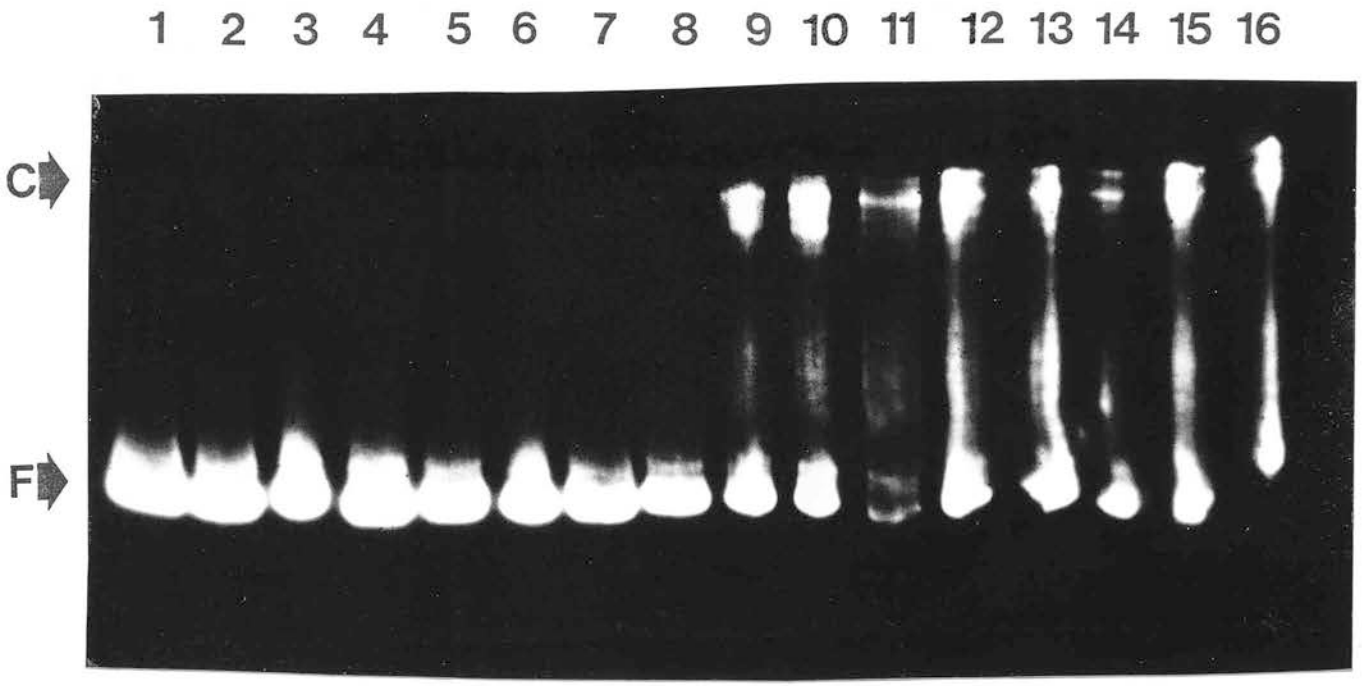


Figure 5. 11. Gel retardation of the 236 bp fragment.

The 236 bp *Taq*I fragment (-46 to +190) was incubated with different protein fractions isolated by heparin-Sepharose chromatography, as in figure 5.5. Sample details as in figure 5.6. F, free DNA; C, protein/DNA complexes.

Lane	:	1	2	3	4	5	6	7	8
Fraction	:	-	2	4	6	8	10	12	14

Lane	:	9	10	11	12	13	14	15	16
Fraction	:	16	17	18	19	20	21	22	23

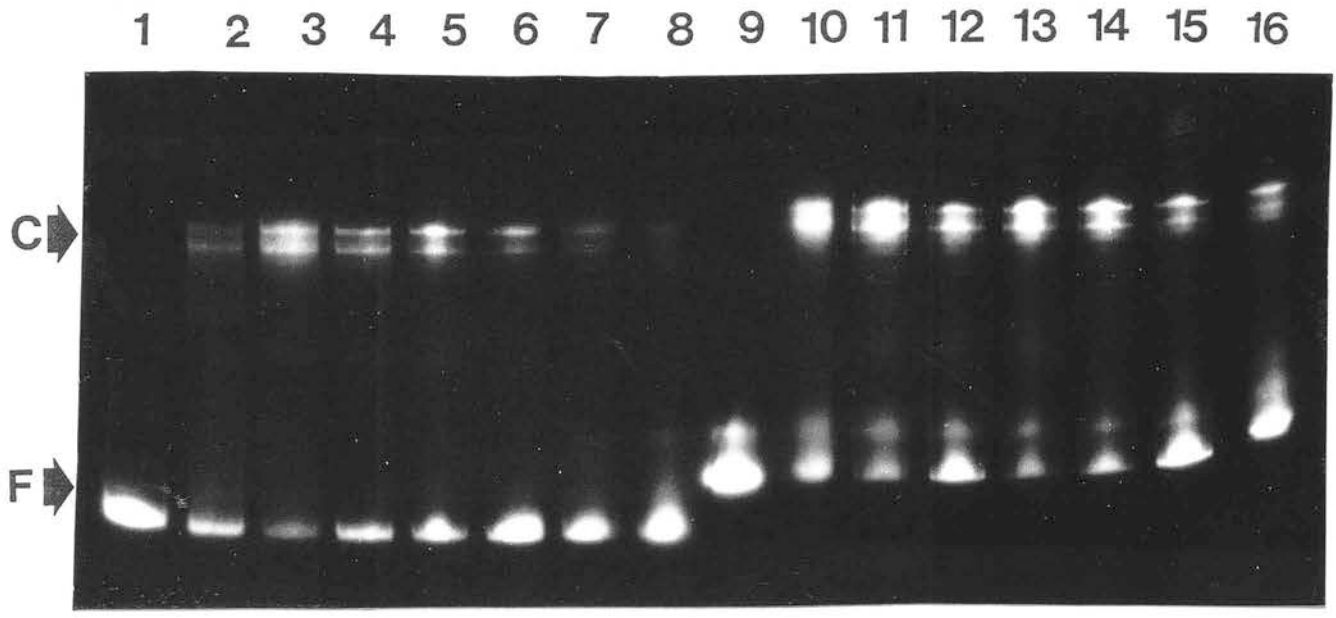


Figure 5. 12. Gel retardation of the 165 and 236 bp *TaqI* fragments.

The 165 bp (lanes 1 to 8) and 236 bp (lanes 9 to 16) *TaqI* fragment were incubated with different protein fractions isolated by heparin-Sepharose chromatography, as in figure 5.5. Sample details as in figure 5.6. F, free DNA; C, protein/DNA complexes.

Lane :	1	2	3	4	5	6	7	8
Fraction :	-	17	18	19	20	21	22	23

Lane :	9	10	11	12	13	14	15	16
Fraction :	-	17	18	19	20	21	22	23

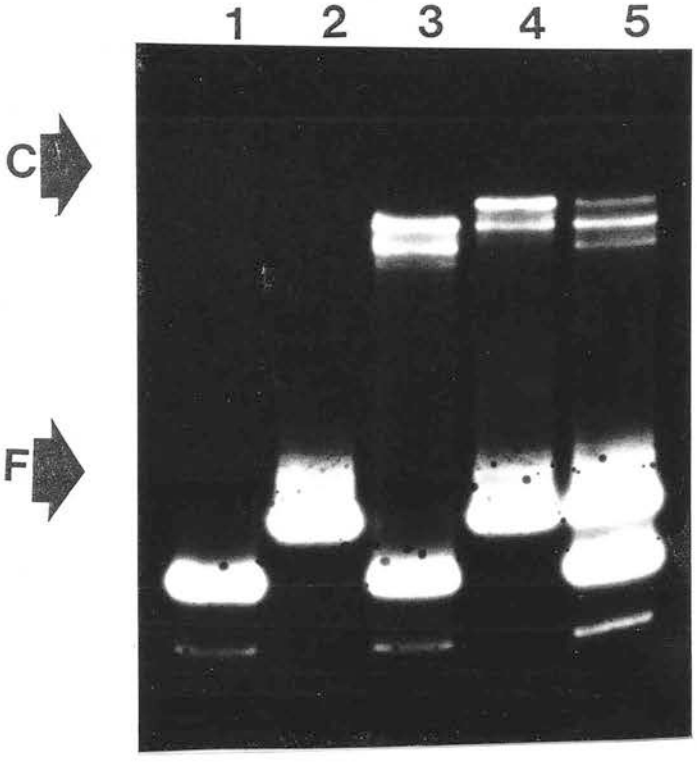


Figure 5. 13. Gel retardation of the 165 and 236 bp *TaqI* fragments mixed together.

The 165 and 236 bp *TaqI* fragments were incubated both separately and together with fraction 19 isolated by heparin-Sepharose chromatography, as in figure 5.5. Sample details as in figure 5.6. F, free DNA; C, protein/DNA complexes.

Lane	:	1	2	3	4	5
Fragment	:	165	240	165	240	165/240
Fraction	:	-	-	19	19	19

5.3.3 Competition experiments to investigate the location of binding sites in the 165 and 236 base pair fragments

A 23 bp synthetic oligonucleotide with the sequence ATTCACGTGAA(T/A)TTCACGTGAAT was constructed in an attempt to investigate the location of binding sites for the protein present in fraction 19. The oligonucleotide, is very similar to the sequence from -282 to -260 within the *LPD* gene and contains two perfect copies of the CP1 binding sequence. The palindromic nature of the molecule allowed it to self anneal. Indeed, analysis of the oligonucleotide by agarose gel electrophoresis revealed it was already predominantly double stranded when obtained from the OWSWEL DNA synthesis service (result not shown). This annealed oligonucleotide was used in competition experiments to establish if increasing amounts of unlabelled double stranded oligonucleotide could compete for the protein binding to the 165 and 236 bp fragments.

The result of the competition experiment are shown in figure 5.14. The annealed oligonucleotide showed a limited ability to compete for the protein. Even when present in quantities equivalent to 100-fold molar excess, with respect to the putative binding site, the double stranded oligonuceotide was unable to prevent the formation of a small amount of labelled DNA/protein complex. The 23 bp fragment was blunt-end ligated and the competition experiment repeated to investigate whether the length of the competing molecule affected its ability to compete for the DNA-binding protein. The results of this experiment are shown in figure 5.15. The ligated double stranded oligonucleotide proved to be a better competitor for the protein binding to the 165 and 236 bp fragments indicating that the increased length of the competing molecule did improve its ability to compete for the DNA-binding protein present. No estimate was made of the length of the ligated double stranded

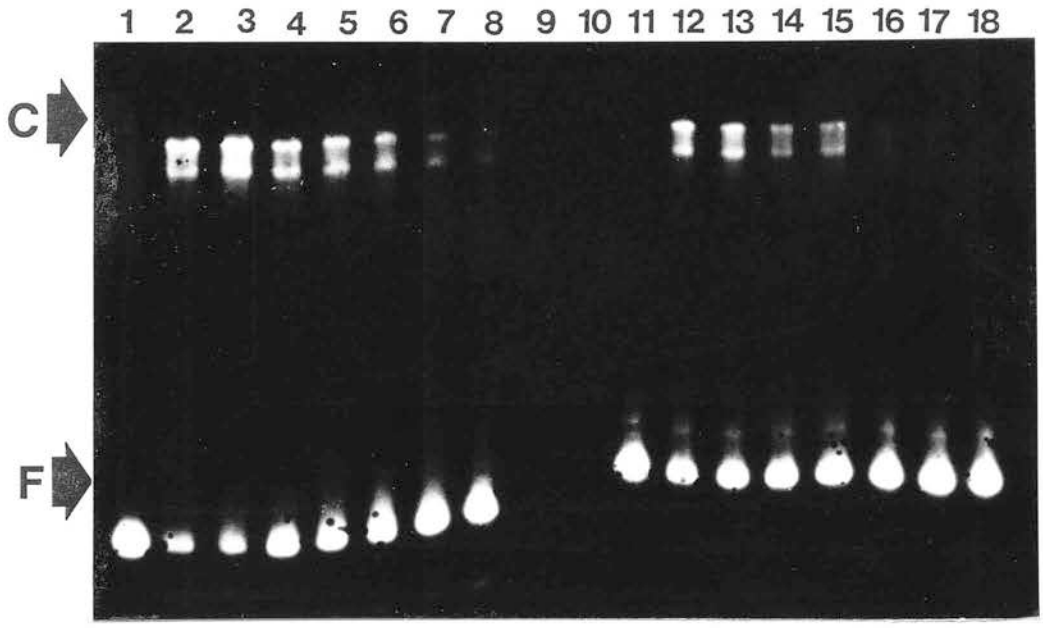


Figure 5. 14. Competition experiment using annealed synthetic oligonucleotide.

Increasing amounts of the oligonucleotide were added to 0.5 ng of the 165 or 236 bp *TaqI* fragment prior to incubation with 1.05 μg of fraction 19 isolated by heparin-Sepharose chromatography, as in figure 5.5. Sample details as in figure 5.6. Lanes 1 and 11 contain no protein fraction. Lanes 1 to 8 contain 165 bp fragment and lanes 11 to 18 contain 240 bp fragment. F, free DNA; C, protein/DNA complexes.

Lane	: 1	2	3	4	5	6	7	8	9
Oligonucleotide (ng):	-	-	0.1	0.5	1.0	2.0	5.0	10	-

Lane	: 10	11	12	13	14	15	16	17	18
Oligonucleotide (ng):	-	-	-	0.1	0.5	1.0	2.0	5.0	10

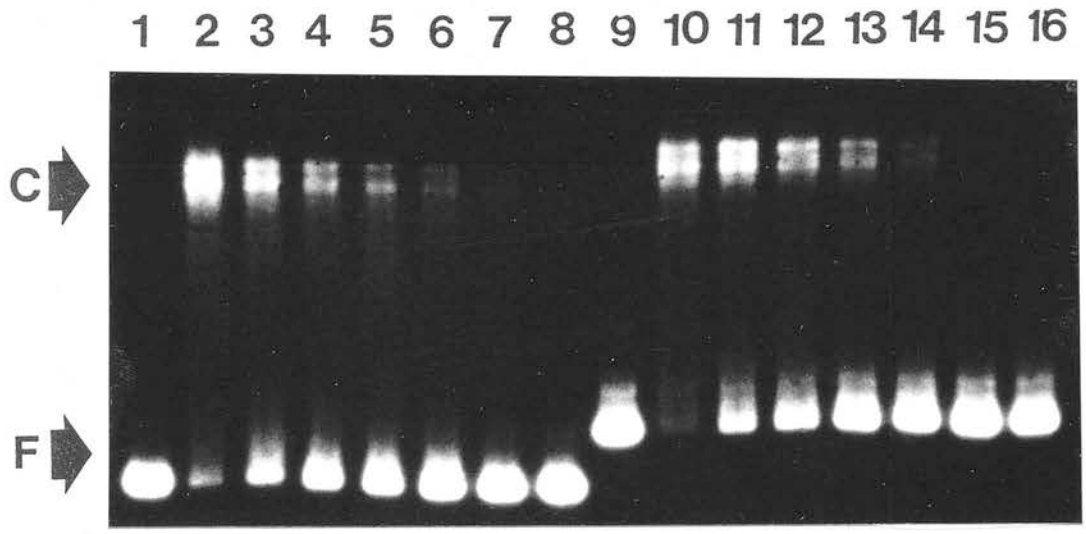


Figure 5. 15. Competition experiment using ligated annealed synthetic oligonucleotide.

Increasing amounts of the oligonucleotide were added to 0.5 ng of 165 or 236 bp *TaqI* fragments prior to incubation with 1.05 μg of fraction 19 isolated by heparin-Sepharose chromatography, as in figure 5.5. Sample details as in figure 5.6. Lanes 1 and 9 contain no protein fraction. Lanes 1 to 8 contain 165 bp fragment and lanes 9 to 16 contain 240 bp fragment. F, free DNA; C, protein/DNA complexes.

Lane	: 1	2	3	4	5	6	7	8
Oligonucleotide (ng):	-	-	0.1	0.5	1.0	2.0	5.0	10

Lane	: 9	10	11	12	13	14	15	16
Oligonucleotide (ng):	-	-	0.1	0.5	1.0	2.0	5.0	10

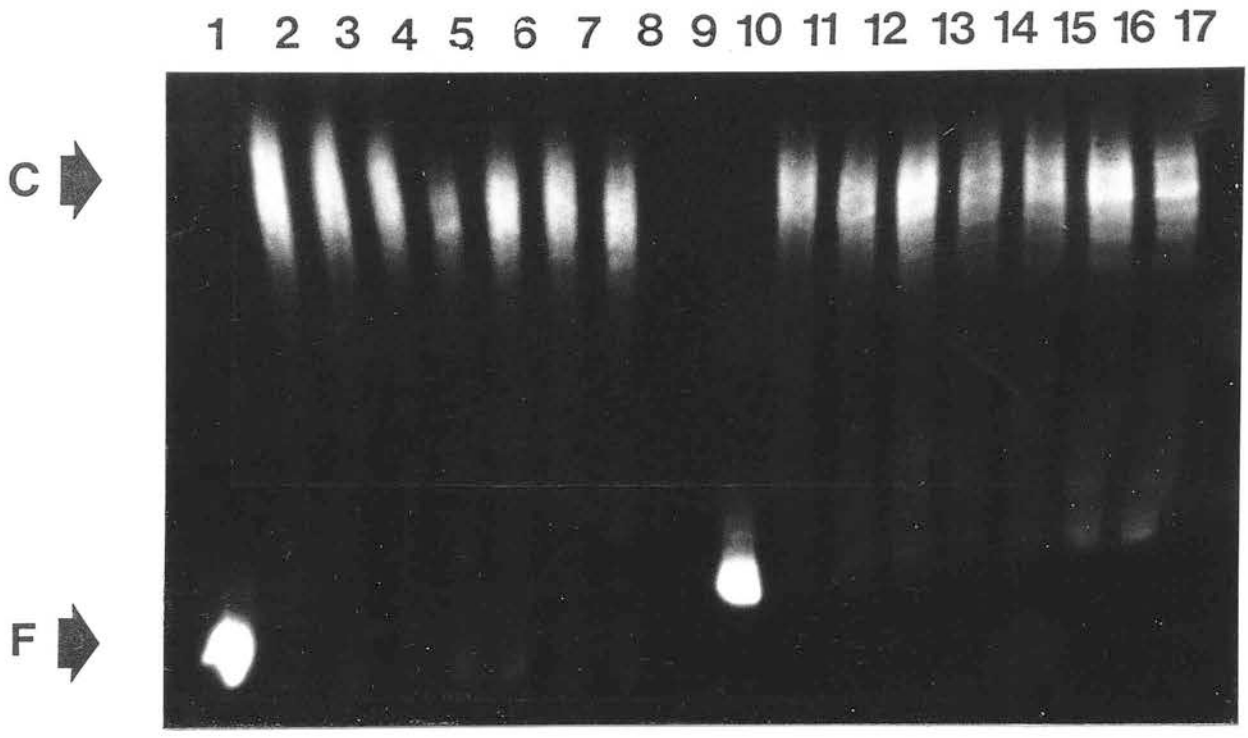


Figure 5. 16. Competition experiment using a pBR322 plasmid containing a *Deinococcus radiodurans* insert.

Increasing amounts of the plasmid DNA were added to 0.5 ng of 165 or 236 bp *TaqI* fragments prior to incubation with 1.05 µg of fraction 19 isolated by heparin-Sepharose chromatography, as in figure 5.5. Sample details as in figure 5.6. Lanes 1 and 10 contain no protein fraction. Lanes 1 to 8 contain 165 bp fragment and lanes 10 to 17 contain 236 bp fragment. Lane 9 is blank. F, free DNA; C, protein/DNA complexes.

Lane	: 1	2	3	4	5	6	7	8	9
Plasmid DNA (ng):	-	-	0.1	0.5	1.0	2.0	5.0	10	-

Lane	: 10	11	12	13	14	15	16	17
Plasmid DNA (ng):	-	-	0.1	0.5	1.0	2.0	5.0	10

oligonucleotides as succesful ligation was seen as a smear of DNA on an agarose gel (results not shown). The competition experiment was repeated using an equivalent amount of a pBR322 plasmid derivative containing a DNA fragment from the bacterium *Deinococcus radiodurans*. The results of this experiment are shown in figure 5.16. No significant competition for the protein binding to the 165 and 236 bp DNA fragments, as assayed by the disappearance of the observed retardation complexes, was detected using this plasmid DNA as a competitor. The significance of these three competition experiments will be discussed below in a critical review of the experiments presented in this chapter.

5.4 CRITICAL ANALYSIS OF THE EXPERIMENTS PERFORMED AND SUGGESTIONS FOR IMPROVEMENT OF EXPERIMENTAL DESIGN

The decision to fractionate a yeast crude extract using heparin-Sepharose affinity chromatography, prior to assaying for specific proteins which interacted with the upstream region of the *LPD* gene by gel retardation, will be discussed in the final chapter as part of a more general discussion of approaches to the analysis of a regulatable promoter. Alternative techniques to assay for DNA-binding proteins will also be presented in chapter 6. This section will therefore be confined to a critical review of the experimental approaches used to determine the specificity of the DNA-binding activities detected, by gel retardation experiments, and to suggestions for alternative approaches to demonstrate the specificity of protein/DNA interactions using gel retardation.

All of the results presented in this chapter demonstrate that the fractions, isolated from crude yeast extract and purified by heparin-Sepharose affinity chromatography, contain proteins capable of retarding the labelled DNA fragments with which they were incubated. The next step

once protein-DNA interactions had been detected was to determine the specificity of these interactions as a wide range of non-specific DNA-binding proteins could be responsible for the observed retardation complexes. The initial technique used to demonstrate the specificity of the protein/DNA interactions involved the use of more than one DNA fragment in a particular gel retardation assay. The assertion here was that if one or more of the fragments included in the assay was not shifted the retardation of the other fragment(s) would represent specific binding. It was also envisaged that such an approach, involving the use of several different restriction digests of the same upstream DNA fragment prior to gel retardation, would help to locate the site of binding of any specific DNA-binding protein.

In the gel retardation assays presented in figures 5.1, 5.2, 5.3 and 5.5 four different sets of restriction fragment, covering different sections of the 1022 bp *SalI/KpnI* region, were used in an attempt to both localize the detected DNA-binding activity and to demonstrate specificity. In all four of the figures presented, however, the retarded fragments represented the largest fragments present in the assay. It is therefore possible that the preferential retardation of the respective fragments in each assay occured simply due to the larger size of these fragments.

This criticism can also be extended to the remainder of the results presented in this chapter. In figures 5.6 to 5.16 *TaqI* DNA fragments were used to assay for DNA-binding proteins present in heparin-Sepharose fractions. Again the only results in which proteins bound preferentially to DNA fragments involved the largest fragments used in this series of experiments (as already discussed earlier the binding activity observed in figures 5.6, 5.7, and 5.8 may be non-specific as retardation activities

were observed with all the fragments tested (results for the 236 and 165 bp fragments not shown)).

The nature of the results obtained in the gel retardation experiments presented in this chapter, which involved the total retardation of larger fragments while smaller fragments remained unaffected, suggest some form of preferential protein/DNA interaction is taking place. Without results demonstrating the retardation of small fragments relative to larger fragments which remain unshifted though no firm conclusions, regarding the specificity of the DNA-binding proteins identified, can be reached from the results presented.

An alternative approach to attempt to demonstrate the specificity of the DNA-binding proteins which retard the 236 and 165 bp fragments involved the use of a 23 bp oligonucleotide. This molecule was designed to be both palindromic (and thus capable of self-annealling) and to contain two copies of the sequence TCACGTGA similar to a region found within both the 165 and 236 bp *TaqI* fragments. A sequence similar to this has also been shown to have a role as a DNA-binding site in other yeast genes (see section 5.3.4). Increasing amounts of the double stranded oligonucleotide were clearly able to compete with the radiolabelled *TaqI* fragments for the DNA-binding protein.

The results presented in figures 5.14 and 5.15 do not however demonstrate the specificity of the DNA-binding protein for the two *TaqI* fragments. This is because the only control used in these competition experiments, shown in figure 5.16, was a pBR322-derived plasmid, containing a *D. radiodurans* insert, used to show a similar amount of unrelated DNA (in terms of mass) did not significantly compete for the DNA-binding protein retarding the 165 and 236 bp fragments. The use of this DNA-species as a control for the results obtained using the double

stranded oligonucleotide is not, however, valid. The two types of competing DNA molecules used differ in both size and DNA conformation. As nothing is known about the parameters affecting the DNA-binding protein which interacts with the 165 and 236 bp fragments, the only valid control would be an oligonucleotide of the same length as that used in figures 5.14 and 5.15 which contained an unrelated nucleotide sequence. Thus the results presented in these three figures are again insufficient as evidence to demonstrate the specificity of proteins binding to DNA fragments upstream of the *LPD* gene.

How then could the identification of specific protein/DNA interactions be demonstrated? The first approach discussed above would have demonstrated specific protein/DNA interactions if a fragment of intermediate size had been retarded. A non-specific binding activity would be expected to retard all of the DNA fragments present and the observation that larger fragments remained unshifted would exclude the possibility that the size of the fragment was responsible for its specific retardation.. The use of different restriction digests to produce different families of DNA fragments could thus be used to investigate the specificity of the DNA-binding proteins identified by gel retardation. If the larger fragments present in each set of digests continue to be retarded it would have to be concluded that the DNA-binding proteins are non-specific.

An alternative approach to demonstrate specificity, representing an extension of the oligonucleotide competition studies discussed above, has been described by Hennighausen and Lubon (1988). Following the retardation of a particular DNA fragment during gel retardation a similar assay is performed in which increasing amounts of the same DNA fragment, which has been end-filled (to avoid any affects 3' or 5' overhangs might

have) but not radiolabelled, is added to the labelled DNA prior to incubation with the DNA-binding protein. The unlabelled fragment should compete for the DNA-binding activity resulting in the disappearance of labelled protein/DNA complex as the amount of unlabelled fragment added is increased. The specificity of the DNA-binding protein for the fragment in question can then be demonstrated if an unrelated fragment of the same length fails to compete for the binding activity to the same degree. Again the importance of the similarity in size and DNA conformation between the two DNA species used as competitors needs to be stressed. The use of a very different DNA species as a control is not valid. This approach could thus be used to examine the specificity of the proteins binding to the 165, 236, and 114 bp *TaqI* fragments. These two methods should clearly determine the specificity of the DNA-binding proteins identified by gel retardation assays in this chapter. In the final chapter a review of the range of other techniques available to both isolate DNA-binding proteins and assay for specific protein/DNA interactions will be presented and related specifically to the *LPD* gene.

The conclusions drawn from the results presented in this chapter, bearing in mind the criticisms presented in this section, are therefore similar to those discussed at the end of chapter 3. A series of experiments intended to investigate aspects of the regulatory apparatus of the *LPD* gene, in this case the binding of specific proteins to the 5' end of the gene, have provided suggestive but not conclusive results. Proteins have been identified which bind to this upstream region but the specificity of these interactions has not been unambiguously demonstrated.

CHAPTER 6 DISCUSSION

The initial observation, presented in this work, was the identification of the *LPD* gene as the structural gene for lipoamide dehydrogenase in *Saccharomyces cerevisiae*. DNA sequence analysis of a 2.7 kb region, part of a 5.5 kb fragment which complemented the *lpd* phenotype, revealed the presence of a 1.5 kb open reading frame. The amino acid sequence encoded by this open reading frame showed strong homology to the primary sequences of lipoamide dehydrogenase from E. coli, *A. vinlandii* and pig heart. Comparison of the DNA sequence representing the 5' flanking region of the *LPD* gene with previously identified sequences, known to have a regulatory role in other yeast genes, revealed several sites of homology suggesting the *LPD* gene might have regulatory *cis*-acting elements upstream of the 1.5 kb open reading frame encoding lipoamide dehydrogenase.

The approaches used to study the regulation of the *LPD* gene have been described and the results obtained discussed in chapters 4 and 5. In this final chapter, therefore, the experimental approaches currently available for the study of a regulatable promoter in yeast will be discussed. This discussion will be related specifically to the *LPD* gene and will thus form a framework for future directions in the study of the regulation of this gene.

6.1 ANALYSIS OF *LPD* EXPRESSION UNDER DIFFERENT CONDITIONS

Before any firm conclusions regarding the elements involved in the regulation of the *LPD* gene can be reached some knowledge of the rates of *LPD* expression under different growth conditions is required. Currently the only information regarding rates of *LPD* expression in yeast during growth under different conditions is the analysis of transcript levels during amino acid starvation conditions described in section 4.2 of this thesis and the data described by Roy and Dawes (1987) which suggest the *LPD* gene is subject to catabolite repression. A much more extensive analysis of rates of *LPD* expession is required prior to the development of experiments to investigate the regulatory apparatus of the *LPD* gene. As discussed in section 4.2 any such analysis must include the use of loading controls to ensure any differences in *LPD* transcript abundance are not due to loading differences.

The rates of *LPD* expression need to be tested under a wide variety of conditions. As discussed in section 4.2 a more thorough analysis of *LPD* transcript levels during amino acid starvation is needed to resolve whether or not lipoamide dehydrogenase is regulated by the general amino acid control network. An analysis of rates of *LPD* transcription during growth on a wide variety of carbon sources, such as glucose, galactose, acetate, pyruvate, glycerol, succinate and ethanol, is also needed. A range of other factors may also be involved in controlling *LPD* expression including cofactors such as $FADH_2$, NADH and lipoamide. The effect of heat shock and the presence or absence of oxygen upon *LPD* transcription could also be examined. The observation that 2-oxoglutarate dehydrogenase activity decreases during sporulation initiation led to the current investigation of lipoamide

dehydrogenase regulation (Dickinson *et al.*, 1986) and it would be interesting to now examine how the rate of *LPD* transription is affected by this developmental process.

Perhaps the most interesting aspect of the lipoamide dehydrogenase enzyme is its role in more than one multienzyme complex. As yet no data are available on the ratio of pyruvate dehydrogenase to 2-oxoglutarate dehydrogenase in yeast during growth on different carbon sources. In *E. coli* the ratio of these two enzymes differs significantly depending on the carbon source (see section 1.3). Investigation of the ratio of the specific activities of these two multienzyme complexes during growth on pyruvate and acetate and the corresponding rate of *LPD* transcription should be examined. Antibodies have been raised against all the subunits of the pyruvate dehydrogenase and 2-oxoglutarate dehydrogenase complexes in yeast (Hunter and Lindsay, 1986) and could be used to monitor protein levels of the two complexes as an alternative to enzyme assays which, for 2-oxoglutarate dehydrogenase particularly, can be difficult to perform reproducibly (Ian Dawes, personal communication).

Combinations of growth conditions also need to be examined. It will be also be particularly interesting to observe how potentially conflicting stimuli affect *LPD* expression. For example growth on glucose might be expected to depress *LPD* expression while amino acid starvation may derepress expression. Thus analysis of *LPD* expression during growth on glucose with amino acid starvation conditions should reveal (if both regulatory mechanisms do affect *LPD* expression) which stimulus has the greater effect.

In addition to an analysis of transcript levels the specific activity of lipoamide dehydrogenase under the conditions discussed above needs to be

investigated. Lipoamide dehydrogenase specific antibodies, used by Roy and Dawes (1987), to study the level of lipoamide dehydrogenase protein present in yeast cells during growth on different carbon sources, could also be used for similar investigations under the conditions discussed here. Measurement of lipoamide dehydrogenase protein levels and specific activity, when combined with transcript analysis data, should help to determine whether post-transcriptional and/or post-translational levels of control are involved in regulating lipoamide dehydrogenase specific activity.

A further approach to those discussed above, and one similar to that used in the study of many other regulatable promoters, involves placing the gene for an easily assayed protein, such as the *lacZ* gene which encodes β-galactosidase, downstream of the 5' region of the *LPD* gene. Such a gene fusion once integrated into the yeast genome could be used to examine the nature of the *LPD* promoter by assaying for β-galactosidase activity under the conditions discussed above. It is important to remember, however, that such an approach, although convenient, represents an artificial situation. Accordingly, data obtained using gene fusions should be backed up subsequently by an analysis of transcript levels.

Finally the location of the 5' and 3' termini of the *LPD* transcript have not yet been determined. The transcription termini represent an important characteristic of any gene and their location needs to be established by both primer extension and S1 nuclease mapping as one of the first steps in any further analysis of the *LPD* gene and its regulation. Once a more comprehensive study of the rate of *LPD* transcription under the conditions discussed above has been completed the elements involved in mediating the regulation of the *LPD* gene can be studied effectively.

6.2 METHODS TO IDENTIFY *CIS*-ACTING ELEMENTS AND *TRANS*-ACTING FACTORS INVOLVED IN THE REGULATION OF THE *LPD* GENE IN YEAST

In this section the different methods that have been developed for the analysis of the elements of regulatable promoters in yeast will be discussed. It should also be noted that the various different yeast regulatory mutants which have now been isolated could also be used to investigate *LPD* regulation. As the control mechanisms involved in the regulation of other yeast genes have been investigated mutants, such as *gcn4* (Hinnebusch and Fink, 1983) and *hap2* (Pinkham and Guarente, 1985), which affect specific control networks have been identified. Once the rate of expression of the *LPD* gene under different conditions has been determined as discussed above the investigation of the effect regulatory mutants such as these have upon *LPD* expression represents a powerful method for studying the control networks involved in the control of this gene and one which should be used in combination with the approaches discussed below.

6.2.1 Isolation of protein fractions containing DNA-binding activities

The most common method used for obtaining proteins which can then be assayed, by a variety of techniques (see below), for their ability to bind specifically to the regulatory regions of yeast genes involves the isolation of a nuclear extract. Most methods for nuclear extract isolation are based upon that of Dignam *et al.* (1983) and involve an initial disruption of the cells followed by wash steps to remove the cytoplasmic contaminants from the nuclei before the salt extraction of nuclear proteins and finally dialysis of the extract to obtain a suitable salt concentration for performing DNA-binding studies. Dignam *et al.* (1983) showed the optimal salt concentration for extraction of nuclear proteins for use in *in vitro* transcription studies was 420 mM NaCl. At this ionic strength, partial lysis of nuclei and a high concentration of total extracted protein is observed. This methods with slight variations has since been used by many workers to isolate DNA-binding proteins involved in the regulation of gene expression from both yeast and higher eukaryotes.

The main advantage of this method is its speed and relative simplicity making it a good first step in the investigation of proteins involved in the regulation of a particular gene. It can also be used to rapidly obtain DNA-binding proteins present in nuclear extracts during growth under many different growth conditions. Disadvantages include the low concentration of specific DNA-binding proteins relative to the total protein concentration. Thus the method used to assay for the presence of a specific protein (see below) must be very sensitive. Also, no separation of nuclear proteins is involved preventing any distinction between proteins which bind to the same site from one another.

In addition to the purification of *trans*-acting factors based upon their subcellular location, a variety of affinity chromatographic approaches have been used to further purify particular DNA-binding proteins. Different substrates have been used to isolate DNA-binding proteins including DNA-cellulose (Alberts and Herrick, 1971), heparin-agarose (Ruet *et al.*, 1984) and specific DNA fragments or oligonucleotides, based upon protein/DNA binding studies (see below) bound to cellulose (Rosenfeld and Kelly, 1986; Kadonaga and Tjian, 1986). Proteins binding to a particular column are subsequently eluted using increasing salt concentrations and the protein fractions obtained assayed for specific interaction using radiolabelled DNA fragments (see below). This approach has two main advantages over nuclear extract isolation. Firstly, the protein(s) of interest can be isolated at a higher concentration of the total protein present than by nuclear extract isolation and, secondly, the separation of DNA-binding proteins into different fractions should allow proteins with the same DNA target site to be distinguished. The main disadvantage of this method, when compared with nuclear extract isolation, is the time required to obtain and assay a series of protein fractions. In addition, protein-protein interactions, which may be important *in vivo*, may be lost due to the separation of proteins into different fractions. The use of the affinity chromatography approach to identify specific *trans*-acting factors is thus perhaps best used as a secondary approach following initial investigations using nuclear extracts.

Heparin-sepharose affinity chromatography of yeast crude extract was used in the experiments described in chapter 5 primarily on the basis of the potential complexity of the *LPD* promoter. It was reasoned that if all of the sites, identified as showing homology to other yeast regulatory sequences, were

binding sites for *trans*-acting factors, results using nuclear extracts containing all of these proteins would be difficult to interpret. In addition, other workers had used this approach successfully as a first step in the isolation of class C transcription factor (Ruet *et al.*, 1984; Huet *et al.*, 1985), and had produced gel retardation assay and DNaseI footprinting results of a very high quality relative to many other reports. With hindsight, the length of time required to isolate and assay the different protein fractions obtained along with the potential importance of protein-protein interactions in gene regulation, make the use of the nuclear extract isolation approach an attractive alternative to the method used and one that should be considered in any further investigation of proteins binding to the upstream region of the *LPD* gene.

6.2.2 Assays for specific protein/DNA interactions

Several methods have been developed to assay for specific protein/DNA interactions. The first technique developed for this purpose was nitro-cellulose filter binding described by Jones and Berg (1966). This assay is based upon the differential binding of proteins and double stranded DNA to nitro-cellulose at a wide range of ionic strengths. Nearly all proteins bind to nitro-cellulose whereas double stranded DNA does not bind above 10 mM NaCl. When labelled DNA fragments are incubated with appropriate dilutions of solutions of DNA-binding proteins and protein/DNA complexes are separated from free DNA by adsorption to nitro-cellulose. After elution from nitro-cellulose filters filter bound DNA can be visualised by gel electrophoresis and autoradiography. Specific protein/DNA interactions are observed as preferentially retained DNA fragments. This specificity can be confirmed by competition experiments (Hennighausen and Lubon, 1988). Unlabelled DNA

fragments, which are otherwise identical to those binding to proteins and thus to the nitrocellulose filters, are used to compete with the labelled DNA molecules for protein binding sites. If a protein/DNA interaction is specific this will be seen by the inability of an unrelated DNA fragment of the same size to compete equally well for the DNA-binding protein. This competition assay is similar to the one described for gel retardation assays in section 5.4. Unspecific DNA-protein interactions can be supressed by the addition of unlabelled non-specific DNA.

The main advantage of this method is that it allows a rapid screening of several DNA fragments (up to 5 kb) simultaneously. The method is thus useful as an initial screen for the presence of specific DNA-binding proteins which bind anywhere within a relatively large region of DNA. It is of less use in determining the exact location to which a protein binds and cannot be used to determine the number of proteins binding to any one particular DNA fragment.

A more recently developed technique which has been used extensively in the study of protein/DNA interactions is gel retardation or mobility shift assay (Garner and Revzin, 1981; Fried and Crothers, 1981). This technique has been used and discussed in chapters 4 and 5 and will thus only be described breifly here. The technique relies upon the retarded mobility of DNA fragments when bound to one or more proteins relative to naked DNA during electrophoresis on native agarose or polyacrylamide gels. Thus radio-labelled DNA fragments can be used to assay for specific protein/DNA interactions. Again non-specific unlabelled DNA such as poly(dI):poly(dC) can be included in the assay to prevent non-specific protein/DNA interactions.

Gel retardation is a rapid simple and very sensitive assay. The major advantage this method has over the nitro-cellulose binding assay is the ability to identify the binding of several proteins to one labelled DNA fragment. Protein/DNA complexes display differing mobilities depending upon the number of proteins bound to a particular DNA fragment. Once again competition experiments (see section 5.4) can be used to demonstrate the specificity of a particular protein/DNA interaction. One disadvantage of this method is the limited size of DNA fragment which can be used. DNA fragments larger than 1 kb are not normally used as they fail to migrate through the gel when complexed with protein.

Gel retardation has also been combined with DNase I footprinting (Galas and Schmitz, 1978) which can be carried out on DNA/protein complexes cut from the polyacrylamide gel (Singh *et al.*, 1986). Again the basic method for DNase I footprinting has been discussed previously in chapters 2 and 4. Briefly footprinting by DNase I involves the incubation of an asymetrically labelled DNA fragment with DNA-binding proteins followed by partial digestion of the DNA with DNase I. The partially digested DNA fragments are then denatured and run on a sequenceing gel. Proteins binding to the DNA block the nicking of DNA strands by steric hinderance of the enzyme. Protein protected regions are seen as regions of missing bands or 'footprints' when compared to digestions in the absence of binding protein. This method has the advantage of allowing the detection of multiple protein binding sites simultaneously. The major disadvantage of this method is its low sensitivity requiring 70 to 80% of the DNA molecules to be complexed with the binding protein(s) to permit detection of a significant difference in band intensity.

A variation on the use of DNase I for footprinting takes advantage of the nuclease activity of the 1,10-phenanthroline-copper ion and the subsequent quenching of this activity using 2,9-dimethyl-1,10-phenanthroline (Kuwabara and Sigman, 1987). Essentially the method is the same as that for DNase I footprinting with the replacement of the enzymic activity with a chemical nuclease. This chemical method has two main advantages over DNase I footprinting. Firstly, the chemical nuclease activity results in a better 'ladder' of DNA fragments as unlike DNase I it nicks double stranded DNA totally independently of DNA sequence. Secondly, the method allows a complete retardation gel to be treated to create nicked DNA prior to excision of a DNA/protein complex. This is useful when a long exposure to identify a DNA/protein complex is necessary as it avoids any pertubation of the footprint by protein denaturation.

Exonuclease III footprinting represents another method which involves detection of DNA/protein complexes by the protection of the DNA from a nuclease (Wu, 1985). This technique however is considerably more sensitive than the DNase I or chemical footprinting methods discussed above. Exonuclease III is a progressive 3' to 5' exonuclease strictly dependent on double stranded DNA. Normally naked DNA is degraded from both 3' ends until two exonuclease molecules collide, releasing two approximately half-sized single stranded DNA fragments (Kornberg, 1980). Thus, if proteins are bound specifically to asymmetrically 5' end-labelled double stranded DNA, exonuclease III digestion stops can be mapped on denaturing urea polyacrylamide gels as new bands which should correspond to the borders of protein protection in DNase I footprinting experiments. Exonuclease III footprinting has a major advantage over DNase I

or chemical footprinting as unlike both these techniques, which are 'negative assays' relying upon differences in band intensity for detection, exonuclease III footprinting requires as little as 1% protection to allow detection of protein/DNA complexes.

All four of the techniques discussed in this section could be used in the further investigation of the *LPD* promoter. As discussed in chapters 4 and 5 gel retardation results require careful use of controls and competition experiments to demonstrate specific protein/DNA interactions. This also applies to nitrocellulose filter binding experiments. Both these techniques could be used to screen further protein extracts prepared using the approaches discussed in the preceeding section. The different footprinting techniques would then represent the next step in localising protein binding sites and determining the number of sites to which any specific protein(s) are binding.

6.2.3 *In vivo* analysis of protein/DNA interactions

An alternative approach to examining the interaction of DNA-binding proteins with specific fragments *in vitro* is to attempt to examine *in vivo* protein/DNA interactions. The technique of *in vivo* footprinting has been used for this purpose. The method is an adaption of the technique of genomic sequencing developed by Church and Gilbert (1984), to analyse CpG residues for methylation of the cytosines in the genome (Giniger *et al.*, 1985). Briefly, whole living cells are reacted with dimethylsulphate (DMS) to partially methylate guanine residues within the nucleus. DNA is then purified, digested with a restriction enzyme of choice and chemically cleaved at positions of methylated guanines with piperidine. The resulting fragments are separated on a denaturing polyacrylamide gel, transferred to a nylon membrane by

electroblotting and covalently crosslinked to the membrane with UV light. In order to visualise the pattern of genomic guanine residues (genomic sequence) of interest, the membrane is then hybridised against a short single stranded radiolabelled DNA probe that abuts the chosen restriction fragment. The intensity of each band of the genomic sequence reflects the reactivity of its corresponding guanine residue towards methylation by DMS. Apparent protection or enhancement of methylation at specific guanine residues (in comparison to a reference pattern obtained by treating naked genomic DNA with DMS) is thought to result from binding of proteins at or close to these residues and formation of hydrophobic pockets due to the close contact of DNA and protein.

The main advantage this technique involves is the analysis of protein interactions with target sequences within the genome and its native chromatin enviroment. As such the technique should be of use in the study of the *LPD* promoter as it represents a form of analysis distinct from that provided by the *in vitro* aproaches discussed above.

6.3 METHODS TO INVESTIGATE THE ROLE OF SPECIFIC DNA SEQUENCES IN GENE REGULATION

The above approaches can be used to isolate DNA binding proteins and to define the location(s) at which proteins bind to specific regions of DNA. They do not however directly address the question of what role, if any, these proteins play in regulating gene expression. Furthermore the methods to isolate and assay specific DNA-binding proteins may often be insufficiently sensitive to identify certain *trans*-acting factors involved in gene regulation. Different techniques have been used to examine how the sequential removal or substitution of regions of DNA with a potential regulatory role affect gene expression. Such

an approach identifies those regions of DNA involved in controlling gene expression and can uncover sites not detected using the methods discussed above.

*Bal*31 exonuclease has been used to create sets of deletions in the 5' and 3' direction from specific restrictions sites in the upstream region of many different yeast genes. Such sets of 5' and 3' deletions can then be ligated together to produce a series of window deletions where each construct contains all of the upstream region of a gene minus a small region, the 'window', which differs slightly from one construct to the next. Such constructs can then be transformed into yeast cells or, preferably, integrated into the yeast genome. An analysis of what effect each deletion has upon transcript levels under different conditions thus provides information about which sequences are important in modulating gene expression. An extension of the window deletion approach, termed linker scanning, involves the substitution of the removed sequence with a linker sequence. Thus without altering the relative spacing of surrounding sequences very specific regions can be altered and any affect on gene expression observed. Finally, the devlopment of site directed mutagenesis now enables the role of specific bases in the regulatory apparatus of a gene to be examined. Once a small region of DNA has been identified as a comtrol element specific base changes can be introduced within this region and the effects upon gene regulation analysed. The investigation of the control elements involved in the regulation of *CYC1* (Forsburg and Guarente, 1988), and HIS3 (Hope and Struhl, 1985; Hill *et al.*, 1986), both of which have been discussed in this thesis, represent examples of the use of the three techniques of window deletion, linker scanning and site directed mutagenesis. Using these

techniques the exact regulatory role of specific bases have been determined demonstrating the power of this type of approach in the study of regulatable promoters.

The types of technique mentioned here are clearly applicable to the analysis of the role of the different sequences in the upstream region of the *LPD* gene showing homology to other yeast regulatory sequences discussed in chapter 3. A small series of window deletions, each involving the removal of approximately 50 bp, could perhaps be used initially to localise those regions involved in gene regulation by monitoring the effect of these deletions upon *LPD* expression. Subsequently finer mapping by linker scanning and site directed mutagenesis could be used to locate these regions more precisely. Such information when combined with results from the type of protein/DNA binding studies discussed above should provide some idea of how the regulatory apparatus of the *LPD* gene is organised and what role any proteins binding specifically to the upstream region of the *LPD* gene have in transcriptional control.

6.4 SUMMARY

The analysis of the *LPD* promoter in yeast requires firstly a detailed and extensive examination of how different growth conditions affect the rate of *LPD* expression. Before this is carried out the role of potential *cis* and *trans*-acting factors in the regulation of the gene cannot be established. Once the pattern of *LPD* expression under different growth conditions has been determined the wide range of techniques which have now been developed to study protein/DNA interactions and to identify sequences involved in regulation of gene

expression, discussed throughout this chapter, can be used to study the *LPD* promoter.

Beyond the analysis of the *LPD* promoter, it would be interesting to investigate whether other yeast genes encoding TCA cycle enzymes share common regulatory sequences involved in modulating co-ordinated expression. In particular the genes encoding the E1 and E2 subunits of pyruvate dehydrogenase and 2-oxoglutarate dehydrogenase might be expected to contain certain *cis*-acting elements which are similar to those involved in the regulation of *LPD* expression. Such an investigation will rely upon the isolation and sequencing of the other yeast genes encoding TCA cycle enzymes.

REFERENCES

References from *Yeast* volumes 2 and 4 refer to plenary abstracts from the special conference issues of this Journal.

Alberts, B. & Herrick, G. (1971). DNA-cellulose chromatography. Methods Enzymol. **21D**, 198-217.

Ashburner, M. & Bonner, J. J. (1979). The induction of gene activity in *Drosophila* by heat shock. *Cell* **17**, 241-254.

Atkinson, D. E. (1968). The energy charge of the adenylate pool as a regulatory parameter. Interaction with feedback modifiers. *Biochem. J.* **7**, 4030-4034.

Barrera, C. R., Namihira, G., Hamilton, L., Munk, P., Eley, M. H., Linn, T. C., & Reed, L. J. (1972). Studies on the subunit structure of the pyruvate dehydrogenase complexes from bovine kidney and heart. *Arch. Biochem. Biophys.* **148**, 343-347.

Bates, D. L., Danson, M. J., Hale, G., Hooper, E. A., & Perham, R. N. (1977). Self-assembly and catalytic activity of the pyruvate dehydrogenase multienzyme complex of *Escherichia coli*. *Nature* **268**, 313-316.

Birnboim, H. C. & Doly, J. (1979). A rapid alkaline extraction procedure for sreening recombinant plasmid DNA. *Nucleic Acids Res.* **7**, 1513-1523.

Bram, R. J. & Kornberg, R. D. (1987). Isolation of a *Saccharomyces cerevisiae* Centromere DNA-binding protein, its human homolog, and its possible role as a transcription factor. *Mol. Cell. Biol.* **7**, 403-409.

Bram, R. J., Lue, N. F., & Kornberg, R. D. (1986). A *GAL* family of upstream activating sequences in yeast: roles in both induction and repression of transcription. *EMBO. J.* **5**, 603-608.

Breathnach, P, Chambon, V. L., Maler, B. A., & Yamamoto, K. R. (1981). Organisation and expression of eukaryotic split genes coding for proteins. *Ann. Rev. Biochem.* **50**, 349-383.

Brent, R. (1985). Repression of transcription in yeast. *Cell* **42**, 3-4.

Broach, J. R. Li, Y., Feldman, J., Jayaram, M., Abraham, J., Nasmyth, K. A., & Hicks, J. B. (1982). Localisation and sequence analysis of yeast origins of DNA replication. *Cold Spring Harbour Symp. Quant Biol.* **47**, 1165-1173.

Browning, K. S., Uhlinger, D. J., & Reed, L. J. (1988). Nucleotide sequence for yeast dihydrolicamide dehydrogenase. *Proc. Natl. Acad. Sci. USA* **85**, 1831-1834.

Buchman, A. R., Williams, J. K., Jasper, R., & Kornberg, R. D. (1988). Two DNA-binding factors recognise specific sequences at silencers, upstream activation sequences, autonomously replicating sequences and telomeres in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **8**, 210-225.

Buratowski, S., Hahn, S., Sharp, P. A., & Guarente, L. (1988). Function of a yeast TATA element-binding protein in a mammalian transcription system. *Nature* **334**, 37-42.

Burton, Z. F., Ortalan, L. G., & Greenblatt, J. (1986). Proteins that bind to RNA polymerase II are required for accurate initiation of transcription at the adenovirus 2 major late promoter. *EMBO J.* **5**, 2923-2930.

Celniker, S. E., Sweder, K., Bailey, J.E., & Campbell, J.L. (1984). Deletion mutations affecting autonomously replicating sequence *ARS1* of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **4**, 2455-2466.

Chandler, P. (1983). DNA sequences bound specifically by Glucocortoid Receptors *in vitro* render a heterologous promoter hormone responsive *in vivo*. *Cell*, **33**, 489-499.

Church, G. M. & Gilbert, W. (1984).Genomic sequencing. *Proc. Natl. Acad. Sci. USA* **81**, 1991-1995.

Cooper, T. G., Street, D., Rai, R., Genbauffe, F., & Yoo, H. S. (1988). Genes and sequences associated with nitrogen catabolite repression. *Yeast* **4**, S323.

Cooper, T. G. & Sumrada, R. A. (1986). Regulation of the *CAR1* gene by a global repressor with target sequences in many yeast genes. *Yeast* **2**, S375.

Craigie, R. & Mizuuchi, K. (1986). Role of DNA topology in Mu transposition. *Cell* **45**, 793-800.

Dagert, M. & Ehrlich, S. D. (1979). Prolonged incubation in calcium chloride improves the competence of *E. coli* cells. *Gene* **6**, 23-28.

Dancis. J. & Levitz, M. (1972). Abnormalities of branched chain amino acid metabolism. In: *The metabolic basis of inherited disease* 3rd ed., edited by J. B. Stanbury, Wyngaarden, & D. S. Fredrickson, pp. 426-439. McGraw-Hill, New York.

De Marucci, O. L., Hunter, A., & Lindsay, J. G. (1985). Low immunogenicity of the common lipoamide dehydrogenese subunit (E3) of mammalian pyruvate dehydrogenese and 2-oxoglutarate dehydrogenese multienzyme complexes. *Biochem. J.* **226**, 509-517.

Delforge, J., Messenguy, F., & Wiame, J. G. (1975). The regulation of arginine biosynthesis in Saccharomyces cerevisae. *Eur. J. Biochem.* **57**, 231-239.

Devereux, J., Haeberli, P., & Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**, 387-395.

Dickinson, J. R., Ambler, R. P., & Dawes, I. W. (1985). Abnormal amino acid metabolism in mutants of *Saccharomyces cerevisiae* affected in the initiation of sporulation. *Eur. J. Biochem.* **148**, 405-406.

Dickinson, J. R., Dawes, I. W., Boyd, A. S. F., & Baxter R. L. (1983). ^{13}C NMR studies of acetate metabolism during sporulation of *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **80**, 5847-5851.

Dickinson, J. R., Roy, D. J., & Dawes, I. W. (1986). A mutation affecting lipoamide dehydrogenese, pyruvate dehydrogenese and 2-oxogluterate dehydrogenese activities in Saccharomyces cerevisiae. *Mol. Gen. Genet.* **204**, 103-107.

Digman, J. D., Lebovitz, R. M. & Roeder, R. G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleis Acids Res.* **11**, 1475-1489.

Dunn, T., Hahn, S., Ogden, S., & Shlief, R. (1984). An operator at -250 bp that is required for repression of *araBAD* operon promoter: Addition of DNA helical turns between the operator and promoter hinders repression. *Proc. Natl. Acad. Sci. USA* **81**, 5017-5020.

Entian, K-D. (1986). Glucose repression: a complex regulatory system in yeast. *Microbiol. Sci.* **3**, 366-370.

Feinberg, A. & Vogelstein, B. (1984). A technique for radiolabelling DNA fragments (addendum). *Anal. Biochem.* **137**, 266-267.

Fink, G. R. & Arndt, K. (1988). New transcription factors required for expression of *HIS4*. *Yeast* **4**, S5.

Finley, R. L. & West, R. W. (1988). *GAL* gene promoters: a battleground for multiple positive and negative regulatory factors. *Yeast* **4**, S396.

Fischer, J. A., Giniger, E., Maniatis, T., & Ptashne, M. GAL4 activates transcription in *Drosophila*. *Nature* **322**, 853-856.

Forsburg, S. L. & Guarente, L. (1988). Mutational analysis of upstream activation sequence 2 of the *CYC1* gene of *Saccharomyces cerevisiae*: a *HAP2-HAP3* responsive site. *Mol. Cell. Biol.* **8**, 647-654.

Fried, M. & Crothers, D. M. (1981). Equilibria and kinetics of lac repressor operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Res.* **9**, 6505-6525.

Galas, D. J & Schmitz, A. (1978). DNAase footprinting: simple method for the detection of protein-DNA binding specificity. *Nucleic Acids Res.* **5**, 3157-3170.

Garland, B. P. (1969). Some kinetic properties of pig heart 2-oxoglutarate dehydrogenase that provide a basis for metabolic control of the enzyme activity and also a stoichiometric assay for Coenzyme A in tissue extracts. *Biochem. J.* **92**, 10C.

Garner, M. M. & Revzin A. (1981). A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to the components of the *E. coli* lactose operon regulatory system. *Nucleic Acids Res.* **9**, 3047-3060.

Gascon, S., Newman, N., & Lampen, O. (1968). Comparative studies of the properties of the purified internal and external invertases from yeast. *J. Biol. Chem.* **243**, 1573-1577.

Gedvilaite, A. A., Sasnauskas, K. V., & Janulaitis, A. A. The yeast *S. cerevisiae ADE2* gene promoter regulatory sites. *Yeast* **4**, S398.

Gineger, E., & Ptashne, M. (1987). Transcription in yeast activated by a putative amphipathic α helix linked to a DNA binding unit. *Nature* **17**, 670-672.

Gineger, E., Varum, S. M., & Ptashne, M. (1985). Specific DNA binding of GAL4. A positive regulatory protein in yeast. *Cell*, **40**, 767-774.

Goedde, H. W. & Keller, W. (1967). Metabolic pathways in maple syrup urine disease. In: *Amino acid metabolism and genetic variation*, edited by W. L. Nyhan, pp. 191-214. McGraw-Hill, New York.

Greenberg, M. L., Myers, P. L., Skrivsky, R. C. & Greer, H. (1986). New positive and negative regulators for general amino acid control of amino acid biosynthesis in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **6**, 1820-1829.

Gribskov, M. & Burgess, R. R. (1986). Sigma factors from *E. coli*, *B. subtilis*, phage SPO1 and phage T4 are homologous proteins. *Nucleic Acids Res.* **14**, 6745-6763.

Griffith, J., Hochschild, A., & Ptashne, M. (1986). DNA loops induced by cooperative binding of λ repressor. *Nature* **322**, 750-752.

Grosschedl, R. & Birnstiel, M. L. (1980). Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants in vivo. *Proc. Natl. Acad. Sci. USA* **77**, 1432-1436.

Guarente, L. (1984). Yeast promoters: positive and negative elements. *Cell* **36**, 799-800.

Guarente, L., Lalonde, B., Gifford, P., & Alani, E. (1984). Distinctly regulated tandem upstream activation sites mediate catabolite repression of the CYC1 gene of Saccharomyces cerevisiae. *Cell* **36**, 503-511.

Guest, J. R. & Creaghan, I. T. (1973). Gene-protein relationships of the α-keto acid dehydrogenese complexes of *Escherichia coli K12*: isolation and characterisation of lipoamide dehydrogenese mutants. *J. Gen. Microbiol.* **75**, 197-210.

Guest, J. R. & Rice, D. W. (1984). Molecular genetic approaches to the study of *E. coli* flavoproteins. In *Flavins and flavoprotiens*, pp. 111-113. Edited by R. C. Bray. Berlin: Walter de Gruyter.

Harashima, S. & Hinnebusch, A. G. (1986). Multiple *GCD* genes required for repression of *GCN4*, a transcriptional activator of general amino acid biosynthetic genes in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **6**, 3990-3998.

von Heijne, G. (1986). Mitochondrial targeting sequences may form amphipilic helices. *EMBO. J.* **5**, 1335-1342.

Hennighausen, L. & Lubon, H. (1987). Interaction of protein with DNA *in vitro*. *Methods Enzymol.* **152**, 721-735.

Hill, D. E., Hope, I. A., Macke, J. P., & Struhl, K. (1986). Saturation mutagenesis of the yeast his3 regulatory site: requirements for transcriptional induction and for binding by GCN4 activator protein. *Science* **234**, 451-457.

Hill, D. E. & Struhl, K. (1986). A rapid method for determining tRNA charging levels *in vivo*: analysis of yeast mutants defective in the general control of amino acid biosynthesis. *Nucleic Acids Res.* **14**, 10045-10051.

Hinnebusch, A. G. (1984). Evidence for translational regulation of the activator of general amino acid control in yeast. *Proc. Natl. Acad. Sci. USA* **81**, 6442-6446.

Hinnebusch, A. G. (1988). Mechanisms of gene regulation in the general control of amino acid biosynthesis in *Saccharomyces cerevisiae*. *Microbiol. Rev.* **52**, 248-273.

Hinnebusch, A. G. & Fink, G. R. (1983). Positive regulation in the general amino acid control of *Saccharomyces cerevisiae*. *Proc. Natl. Acad.Sci. USA* **80**, 5374-5378.

Hinnebusch, A. G., Lucchini, G., & Fink, G. R. (1985). A synthetic HIS4 regulatory element confers general amino acid control on the cytochrome *c* gene (*CYC1*) of yeast. *Proc. Natl. Acad. Sci. USA* **82**, 498-502.

Hodgson, J. A., De Marcucci, O. G., & Lindsay, J. G. (1986). Lipoic acid is the site of sustrate-dependent acetylation of the component X in ox heart pyruvate dehydrogenase multienzyme complex. *Eur. J. Biochem.* **158**, 595-600.

Hope, I. A. & Struhl, K. (1985). GCN4 protein, synthesised *in vitro*, binds *HIS3* regulatory sequences: implications for general control of amino acid biosynthetic genes in yeast. *Cell* **43**, 177-188.

Hope, I. A. & Struhl, K. (1986). Functional dissection of a eukaryotic *trans*-acting protein, GCN4, in yeast. *Cell*, **46**, 885-894.

Horchschild, A. H. & Ptashne, M. (1986). Cooperative binding of λ repressors to sites separated by intergral turns of the DNA helix. *Cell* **44**, 681-687.

Huet, J., Cottrell, P., Cool, M., Vignais, M. L., Thiele, D., Marck, C., Buhler, J. M., Sentenac, A., & Fromageot, P. (1985). A general upstream binding factor for genes of the yeast translational apparatus. *EMBO. J.* **4**, 3539-3547.

Hunter, A. & Lindsay, J. G. (1986). Immunological and biosynthetic studies on the mammalian 2-oxogluterate dehydrogenese multienzyme complex. *Eur. J. Biochem.* **155**, 103-109.

Jones, E. W. & Fink, G. R. (1982) in *The Molecular Biology of the Yeast Saccharomyces. Life Cycle and Inheritance*, eds. Strathern, J. N., Jones, E. W., & Broach, J. R. (Cold Spring Harbour Laboratory, Cold Spring Harbour, NY), pp. 181-299.

Jones, O. W. & Berg, P. (1966). Studies on the binding of RNA polymerase to polynucleotides. *J. Mol. Biol.* **22**, 199-209.

Kadonaga, J. T. & Tjian, R. (1986). Affinity purification of sequence-specific DNA binding proteins. *Proc. Natl. Sci. USA* **83**, 5889-5893.

Kakidina, H., & Ptashne, M. (1988). GAL4 activates gene expression in mammalian cells. *Cell*, **52**, 161-167.

Keegan, L., Gill, G., & Ptashne, M. (1986). Separation of DNA binding from the transcription activating function of a eukaryotic regulatory protein. *Science* **231**, 699-703.

Klopotowski, T. & Wiater, A. (1965). Synergism of aminotriazole and phosphate on the inhibition of yeast imidazoleglycerol phosphate dehydratase. *Arch. Biochem. Biophys.* **112**, 562-566.

Kochi, H. & Kikuchi, G. (1976). Mechanism of reversible glycine cleavage reaction in *Arthrobacter globiformis*; function of lipoic acid in the cleavage and synthesis of glycine. *Arch. Biochem. Biophys.* **173**, 71-81.

Kornberg, A. (1980). In *DNA replication*. Freeman and Co. San Fransisco.

Kozak, M. (1981). Possible role of flanking nucleotides in recognition of the AUG initiation codon by eukaryotic ribosomes. *Nucleic Acids Res.* **9**, 5233-5252.

Krasnow, M. A., Matzuk, M. M., Dungan, J. M., Benjamin, H. W., & Cozzarelli, N. R. (1983). In *Mechanisms of DNA replication and recombination*. ed. Cozzarelli, N. R. 637-659.

Kuwabara, M. D. & Sigman, D. S. (1987). Footprinting DNA-protein complexes in situ following gel retardation assays using 1, 10-phenanthroline-copper ion: *E. coli* RNA polymerase-lac promoter complexes. *Biochemistry* **26**, 7234-7238.

Lahti, R., Glumoff, V., Heimo, H., & Mantsala, P. (1988). On the regulation of yeast *ADE4* gene. Deletion analysis of the 5' flanking region. *Yeast* **4**, S405.

Linn, T. C., Pettit, F. H., & Reed, L. J. (1969). α-ketoacid D complexes, X regulation of the pyruvate dehydrogenase complex from beef kidney mitochondria by phoshorylation and dephosphorylation. *Proc. Natl. Acad. Sci. USA.* **62**, 234-241.

Ma, J., Prazibilla, E., Hu, J., Bogorad, L., & Ptashne, M. (1988). Yeast activators stimulate plant gene expression. *Nature* **334**, 631-633.

Ma, J. & Ptashne, M. (1987). Deletion analysis of GAL4 defined two transcriptional activating segments. *Cell*, **48**, 847-853.

Magasinik, B. (1961). Catabolite Repression. *Cold Spring Harbour Symp. Quant. Biol.* **26**, 249-262.

Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982). *Molecular Cloning: a Laboratory Manual.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Martin, K., Huo, L., & Schlief, R. (1986). The DNA loop model for *ara* repression: AraC protein occupies the proposed loop site *in vivo* and repression negative mutants lie in these same sites. *Proc. Natl. Acad. Sci. USA* **83**, 3654-3658.

Mellor, J., Dobson, M. J., Kingsman, A. J., & Kingsman, S. M. (1987). A transcriptional activator is located in the coding region of the yeast *PGK* gene. *Nucleic acids Res* **15**, 6243-6259.

Mellor, J., Rathjen, J., Heath, J. K., & Dowell, S. J. The short sequence CAC(G/A)TGA has multifunctional properties in yeast and mammalian cells: roles in initiation, repression and basal level enhancement of transcription. (in press)

Messenguy, F. & Scherens, B. (1988). The general amino acid control is part of the stress response of the yeast cell: identification of mutations affecting thermotolerance and general control. *Yeast* **4**, S345.

Messing, J. (1983). New M13 vectors for cloning. *Methods in enzymology* **101**, 20-78.

Miles, J. R., & Guest, J. R. (1987). Genetic aspects of the citric acid cycle of *E. coli*. *Biochem. Soc. Symp.* **54**, 45-65.

Miller, J., McLachlan, A. D.,& Klug, A. (1985). Repetative zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO J.* **4**, 1609-1614.

Mueller, P. P. & Hinnebusch, A. G. (1986). Multiple upstream AUG codons mediate translational control of *GCN4*. *Cell* **45**, 201-207.

Mukherjee, B. B., Matthews, J., Horney, O. L. & Reed, L. J. (1965). Resolution and reconstruction of *Escherichia coli* α-ketogluterate dehydrogenese complex. *J. Biol. Chem.* **240**, 2268-2269.

Mvasnikov, A. N. & Smirnov, M. N. (1988). General amino acid control system may be involved in regulation of purine biosynthesis in yeast. *Yeast* **4**, S412.

Myers, P. L., Skrivsky, R. C., Greenberg, M. L. & Greer, H. (1986). Negative regualtory gene for general amino acid biosynthesis in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **6**, 3150-3155.

Niederberger, P., Miozzari, G., & Huetter, R. (1981). Biological role of the general control of amino acid biosynthesis in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **1**, 584-593.

Ogden, J. E., Stanway, C., Kim, S., Mellor, J., Kingsman, A. J., & Kingsman, S. M. (1986). Efficient expression of the *Saccharomyces cerevisiae PGK* gene depends on an upstream activation sequence but does not require TATA sequences. *Mol. Cell. Biol.* **6**, 4335-4343.

Oshima, Y. (1981) in *The Molecular Biology of the Yeast Saccharomyces. Life Cycle and Inheritance*, eds. Strathern, J. N., Jones, E. W., & Broach, J. R. (Cold Spring Harbour Laboratory, Cold Spring Harbour, NY), pp. 159-180.

Pabo, C. O., & Sauer, R. T. (1984). Protein-DNA recognition. *Ann Rev. Biochem.* **53**, 293-321.

Parker, M. G. & Weitzmann, P. D. J. (1972). The regulation of Acinobacter species α-oxoglutarate dehydrogenase complex. *Biochem. J.* **130**, 39P.

Pelham, H. (1985). Activation of heat shock genes in eukaryotes. *Trends in Genet.* **1**, 31-35.

Penn, M. D., Thireos, G., & Greer, H. (1984). Temporal analysis of general control of amino acid biosynthesis in *Saccharomyces cerevisiae*: role of positive regulatory genes in initiation and maintenance of mRNA derepression. *Mol. Cell. Biol.* **4**, 520-528.

Pfeifer, K., Arcangioli, B., & Guarente, L. (1987). Yeast HAP1 activator competes with the factor RC2 for binding to the upstream activator site UAS1 of the *CYC1* gene. Cell **49**, 9-18.

Pinkham, J. L. & Guarente, L. (1985). Cloning and molecular analysis of the *HAP2* locus, a global regulator of respiritory genes in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **5**, 3410-3416.

Plon, S. & Wang, J. (1986). Transcription of the human β-globin gene is stimulated by an SV40 enhancer to which it is physically linked but topologically uncoupled. *Cell* **45**, 575-580.

Pons, G., Raefsky-Estrin, C., Carothers, D. J., Pepin, R. A., Javed, A. A., Jesse, B. W., Ganapathi, M. K., Samols, D., & Patel, M. S. (1988). Cloning and cDNA sequence of the dihydrolipoamide dehydrogenase component of human α ketoacid dehydrogenase complexes. *Proc. Natl. Acad. Sci. USA* **85**, 1422-1426.

Proudfoot, N. J. & Brownlee, G. G. (1976). 3' non-coding region sequences in eukaryotic messenger RNA. *Nature* **263**, 211-214.

Ptashne, M. (1986). Gene regulation by proteins acting nearby and at a distance. *Nature* **322**, 697-701.

Purvis, I. J., Loughlin, L., Bettany, A. J. E., & Brown, A. J. P. (1987). Translation and stability of an *E. coli* β-galactosidase mRNA expression under the control of the pyruvate kinase sequence of *Saccharomyces cerevisiae*. Nucleic acids Res **15**, 7963-7974.

Reed, L. J. (1969). In *Current Topics in Cellular Regulation*. Horecker, B. L. and Stadtman, E. R. Ed., Vol 1. Academic Press, New York, N. Y.

Reed, L. J. (1974). Multienzyme complexes. *Acc. Chem. Res.* **7**, 40-46.

Robinson, J. R., Klein, S. M., & Sagers, R. D. (1973). Glycine metabolism. Lipoic acid as the prosthetic group in the electron transfer protein P2 from *Peptococcus glycinophilus*. *J. Biol. Chem.* **248**, 5319-5323.

Rosenfeld, P. J.& Kelly, T. J. (1986). Purification of nuclear factor I by DNA recognition site affinity chromatography. *J. Biol. Chem.* **261**, 1398-1408.

Ross, J., Reid, G. A., & Dawes, I. W. (1988). The nucleotide sequence of the *LPD1* gene encoding lipoamide dehydrogenase in *Saccharomyces cerevisiae*: comparison between eukaryotic and prokaryotic sequences for related enzymes and identification of potential upstream control sites. *J. Gen Micro.* **134**, 1131-1139.

Roy, D. J. & Dawes, I. W. (1986). Cloning and characterization of the gene lipoamide dehydrogenase in *Saccharomyces cerevisiae*. *J. Gen. Microbiol.* **133**, 925-933.

Ruet, A., Camier, S., Smagowicz, W., Sentenac, A., & Fromageot, P. (1984). Isolation of a class C transcription factor which forms a stable complex with tRNA genes. *EMBO J.* **3**, 343-350.

Sanger, F., Nicklen, S., & Coulsen, A. R. (1977). DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.

Sawadogo, M. & Roeder, R. G. (1985). Factors involved in specific transcription by human RNA polymerase II: analysis by a rapid and quantitative *in vitro* assay. *Proc. Natl. Acad. Sci. USA* 82, 4394-4398.

Schlesinger, M. J., Aliperti, G., & Kelley, P. M. (1982). The response of cells to heat shock. *Trends Biochem. Sci.* 7, 222-225.

Schurch, A., Miozzari, J., & Hutter, R. (1974). Regulation of tryptophan biosynthesis in *Saccharomyces cerevisiae*: Mode of action of 5-methyl-tryptophan and 5-methyl-tryptophan-sensitive mutants. *J. Bact.* **117**, 1131-1140.

Schwartz, E. R. & Reed, L. J. (1970). Regulation of the activity of pyruvate dehydrogenase complex in *E. coli*. *Biochem.* **9**, 1434-1439.

Singh, H., Sen, R., Baltimore, D. & Sharp, P. A. (1986). A nuclear factor that binds to a conserved sequence motif in transcriptional control elements of immunoglobin genes. *Nature* **319**, 154-158.

Smith, H. O. (1980). Recovery of DNA from gels. *Methods Enzymol.* **65**, 371-380.

Spencer, M. E. & Guest, J. R. (1985). Transcriptural analysis of the sucAB aceEF and lpd genes of E. coli. *Mol. Gen. Genet.* **200**, 145-154.

St. John, M. & Davis, R. W. (1984). Sequences that regulate the divergent GAL1-GAL10 promoter in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **4**, 1440-1448.

Stephens, P. E., Lewis, H. M., Darlison, M. G., & Guest, J. R. (1983). Nucleotide sequence of the lipoamide dehydrogenese gene of *Escherichia coli K12*. *Eur. J. Biochem.* **135**, 519-527.

Struhl, K. (1982). Regulatory sites for *HIS3* expression in yeast. *Nature* **300**, 284.

Struhl, K. (1985). Naturally occuring poly(dA-dT) sequences are upstream promoter elements for constitutive transcription in yeast. *Proc. Natl. Acad. Sci. USA* **82**, 8419-8423.

Struhl, K. (1986). Constitutive and inducible *Saccharomyces cerevisiae* promoters: Evidence for two distinct molecular mechanisms. *Mol. Cell. Biol.* **6**, 3847-3853.

Struhl, K. (1987). Promoters, activator proteins, and the mechanism of transcriptional initiation in yeast. *Cell* **49**, 295-297.

Struhl, K. (1988). The JUN oncoprotein, a vertebrate transcription factor activates transcription in yeast. Nature **332**, 647-649.

Subik, J., Kolarov, J., & Lachowicz, T. M. (1972). A mutant of *Saccharomyces cerevisiae* lacking α-ketoglutarate dehydrogenese activity. **27**, 81-84.

Tlsty, T. (1982). In *Gene Amplification*. Schimker Ed. Cold Spring Harbour publication.

Vezinhet, F., Kinnaird, J., & Dawes, I. W. (1979). The physiology of mutants derepressed for sporulation in *Saccharomyces cerevisiae*. *J. Gen. Microbiol.* **115**, 391-402.

Webster, N., Jin, J. R., Green, S., Hollis, M, & Chambon, P. (1988). Yeast UAS_G is a transcriptional enhancer in human HeLa cells in the presence of the GAL4 *trans*-activator. *Cell* **52**, 169-178.

Weir-Thompson, E. M., & Dawes, I. W. (1984). Developmental changes in translatable RNA species associated with meiosis and spore formation in *Saccharomyces cerevisiae*. Mol Cel. Biol. **4**, 695-702.

Westphal, A. H. & de Kok, A. (1988). Lipoamide dehydrogenase from *Azotobacter vinelandii*. Molecular cloning, organization and sequence analysis of the gene. *Eur. J. Biochem.* **172**, 299-305.

Wieland, O. H. (1983). The mammalian pyruvate dehydrogenese complex: structure and regulation. *Physiol. Biochem. Pharmacol.* **96**, 123-170.

van Wijk, R., Ouwehand, T., van de Bos T., & Koningsberger, V. (1969). Induction and catabolite repression of α-glucosidase synthesis in protoplasts of *Saccharomyces cerevisiae*, Biochim. Biophys. Acta **186**, 178-191.

Williams, C. H. Jr. (1976). Flavin-containing dehydrogenases. In *The Enzymes*, Vol. 13, 3rd ed., edited by P. D. Boyer, pp. 89-173. Academic Press, New York, San Francisco, and London.

Williams, C. H. Jr, Arscott, L. D., & Schultz, G. E. (1982). Amino acid sequences of tryptic peptides isolated from pig heart tissue. *Proc. Natl. Acad. Sci. USA* **79**, 2199-2207.

Williams, C. H. Jr, Arscott, L. D., & Swenson, P. (1984). Active site chemical modification and sequencing of flavoproteins. In *Flavins and Flavoproteins*, pp. 95-109. Edited by R. C. Bray. Berlin: Walter de Gruyter.

Winter, G. & Fields, S. (1980). Cloning of influenza cDNA into M13: the sequence of the RNA segment encoding the A/PR/8/34 matrix protein. *Nucleic Acids Res.* **8**, 1965-1974.

Wittenberger, C. L. & Haaf, A. S. (1964). Lactate-degrading system in *Butyribacterium rettgeri* subject to glucose repression. *J. Bacteriol.*, **88**, 896-903.

Wu, C. (1985). An exonuclease protection assay reveals heat-shock element and TATA box DNA-binding proteins in crude nuclear extracts. *Nature* **317**, 84-85.

Zaret, K. S. & Sherman, F. (1982). DNA sequences required for efficient transcription termination in yeast. *Cell* **28**, 563-573.

Journal of General Microbiology (1988), **134**, 1131–1139. *Printed in Great Britain*

The Nucleotide Sequence of the *LPD1* Gene Encoding Lipoamide Dehydrogenase in *Saccharomyces cerevisiae*: Comparison between Eukaryotic and Prokaryotic Sequences for Related Enzymes and Identification of Potential Upstream Control Sites

By JOE ROSS,* GRAEME A. REID AND IAN W. DAWES

Department of Microbiology, University of Edinburgh, West Mains Road, Edinburgh EH9 3JG, UK

(*Received 30 November 1987; revised 29 January 1988*)

The complete nucleotide sequence of the *LPD1* gene, which encodes the lipoamide dehydrogenase component (E3) of the pyruvate dehydrogenase and 2-oxoglutarate dehydrogenase multienzyme complexes of *Saccharomyces cerevisiae*, has been established. The flanking region 5′ to the *LPD1* gene contains DNA sequences which show homology to known control sites found upstream of other yeast genes. The primary structure of the protein, determined from the DNA sequence, shows strong homology to a group of flavoproteins including *Escherichia coli* lipoamide dehydrogenase and pig heart lipoamide dehydrogenase. The amino acid sequence also reveals the presence of a potential targeting sequence at its N-terminus which may facilitate transport to and entry into mitochondria.

INTRODUCTION

Lipoamide dehydrogenase (EC 1.8.1.4) is a component of the multienzyme complexes pyruvate dehydrogenase and 2-oxoglutarate dehydrogenase, which catalyse the oxidative decarboxylation of pyruvate and 2-oxoglutarate to acetyl-CoA and succinyl-CoA respectively (Reed, 1974). Both complexes contain multiple copies of three component enzymes: pyruvate dehydrogenase (EC 1.2.4.1) or 2-oxoglutarate dehydrogenase (EC 1.2.4.2) (E1); dihydrolipoamide acetyltransferase (EC 2.3.1.12) or dihydrolipoamide succinyltransferase (EC 2.3.1.61) (E2) and lipoamide dehydrogenase (E3). The E1 and E2 components are specific to their respective complexes whereas lipoamide dehydrogenase has been shown to be functionally interchangeable between both (Mukherjee *et al.*, 1965), and in *Escherichia coli* and *Saccharomyces cerevisiae* is encoded by a single gene (Guest & Creaghan, 1973; Dickinson *et al.*, 1986).

In higher eukaryotes lipoamide dehydrogenase has an additional role in the multienzyme complexes which specificially catalyse the oxidative decarboxylation of branched-chain 2-oxoacids derived from leucine, valine and isoleucine by transamination (Lawson *et al.*, 1983). The reversible oxidative decarboxylation of glycine has also been shown to involve lipoamide dehydrogenase in the aerobic bacterium *Arthrobacter globiformis* (Kochi & Kikuchi, 1976), in the anaerobe *Peptococcus glycinophilus* (Robinson *et al.*, 1973) and in rat liver mitochondria (Kochi & Kikuchi, 1976).

Regulation of the synthesis of the E1, E2 and E3 components of pyruvate dehydrogenase and 2-oxoglutarate dehydrogenase in *E. coli* has been studied extensively (Guest & Rice, 1984; Spencer & Guest, 1985), and it has been shown that the *E* and *F* genes of the *ace* operon and the *A* and *B* genes of the *suc* operon encode the E1 and E2 subunits of the pyruvate dehydrogenase and 2-oxoglutarate dehydrogenase complexes respectively. Lipoamide dehydrogenase is encoded by a single gene, *lpd*, which is linked to the *ace* operon but can operate under the control of its own promoter. Transcription from *sucAB*, *aceEF* and *lpd* genes shows differential

regulation in response to the nature of the carbon source, thereby ensuring the necessary ratio of subunit components when pyruvate and/or 2-oxoglutarate dehydrogenase enzyme complexes are required. The absence of operons in eukaryotic organisms dictates that the regulation of synthesis of the E1, E2 and E3 components in yeast must differ from the system found in *E. coli*.

The role of lipoamide dehydrogenase in at least two multienzyme complexes raises several interesting questions concerning the regulation of synthesis and control of distribution of an enzyme which functions within two different complexes. The enzyme is encoded in the nucleus but must be transported to its site of action in the mitochondrion. Biochemical studies of the initiation of sporulation in *S. cerevisiae* have shown that the regulation of 2-oxoglutarate dehydrogenase activity may be of crucial importance in the switch of cells from vegetative growth via mitosis to the developmental processes of meiosis and sporulation (Dickinson *et al.*, 1986). In *S. cerevisiae* the enzyme has also been shown to be subject to catabolite repression (Roy & Dawes, 1987).

Determination of the DNA sequence for the genes which encode the E1, E2 and E3 components of both pyruvate and 2-oxoglutarate dehydrogenase is necessary in order to study how their syntheses are regulated at the molecular level.

METHODS

Sources of DNA. The 3·7 kb *Xho*I fragment which contains the *LPD1* gene was obtained from pGP-R1 (Roy & Dawes, 1987) for 'shotgun' cloning into M13 vectors. Large-scale plasmid and phage RF2 DNA isolation was performed by alkaline-SDS lysis; the DNA was purified by centrifugation in CsCl/ethidium bromide gradients (Maniatis *et al.*, 1982). Restriction fragments were separated by electrophoresis in agarose gels and extracted by electro-elution onto dialysis membrane (Smith, 1980).

Cloning in the M13-based vector. The 3·7 kb *Xho*I fragment was digested with three restriction enzymes for shotgun cloning into M13mp18 (Messing, 1983). *Taq*I and *Msp*I fragments were cloned into the *Acc*I site of M13mp18, and *Sau*3AI fragments were cloned into the *Bam*HI site of the same vector. In addition, the 1·45 kb *Xba*I fragment was cloned into the *Xba*I site of M13mp18. Transfection of *E. coli* JM101 [Δ(*lac–pro*) *supE thi*/*F' traD36 proAB lacI*^{q} ZΔ*M15*] was performed according to published procedures (Winter & Fields, 1980).

Nucleotide sequence analysis. Single-stranded M13 DNA templates were prepared and sequenced by the dideoxy chain-termination method using a 17-nucleotide synthetic primer (Sanger *et al.*, 1977). Two 15-nucleotide synthetic primers, based on information obtained from the *LPD1* sequence, were also used. The nucleotide sequence was compiled and analysed using the University of Wisconsin Genetics Computer Group programs (Devereux *et al.*, 1984).

Materials. Restriction enzymes were purchased from Bethesda Research Laboratories, Boehringer, and Northumbria Biologicals. Phage T4 DNA ligase was from either Boehringer or Northumbria Biologicals. The 17-nucleotide synthetic primer was obtained from Pharmacia; the 15-nucleotide synthetic primers were purchased from OSWEL DNA Synthesis Service, Chemistry Department, Edinburgh University. DNA polymerase (Klenow fragment), dideoxynucleotides and deoxynucleotides were obtained from Boehringer, [α-^{32}P]dATP from Amersham, and reverse transcriptase from Northumbria Biologicals.

RESULTS AND DISCUSSION

The nucleotide sequence

Based on deletion analysis of the plasmid pGP1 (Roy & Dawes, 1987), a YEp13-based vector containing a 5·5 kb yeast DNA insert carrying the *LPD1* gene, a 3·7 kb *Xho*I fragment was chosen for 'shotgun' cloning analysis. A region of 2701 bp has been sequenced. All of the sequence was obtained from at least two independent clones, it was fully overlapping and was, with the exception of a 0·1 kb region 0·8 kb upstream of the coding region, derived from both DNA strands. The program MAP was used to identify the coding region representing the *LPD1* gene. One large open reading frame of 1·5 kb was found, which is consistent with the known size of the yeast lipoamide dehydrogenase subunit polypeptide (Wieland, 1983) and which exhibits a consistently high score with respect to preferred codon usage in *S. cerevisiae* determined using the programs CODONFREQUENCY and CODONPREFERENCE (data not shown). The complete and unambiguous sequence of the *LPD1* gene and the primary structure of lipoamide dehydrogenase, translated from the DNA sequence, are presented in Fig. 1.

In the region 5′-distal to the *LPD1* sequence are two potential open reading frames: one (at −571), which would be read in the opposite sense, is capable of encoding a polypeptide of 77 residues; the other (sequence not complete, terminating at −858) could encode a polypeptide of at least 76 amino acids. A search of the NBRF protein data base showed no strong homology between the two predicted amino acid sequences and any known protein.

Primary structure of yeast lipoamide dehydrogenase

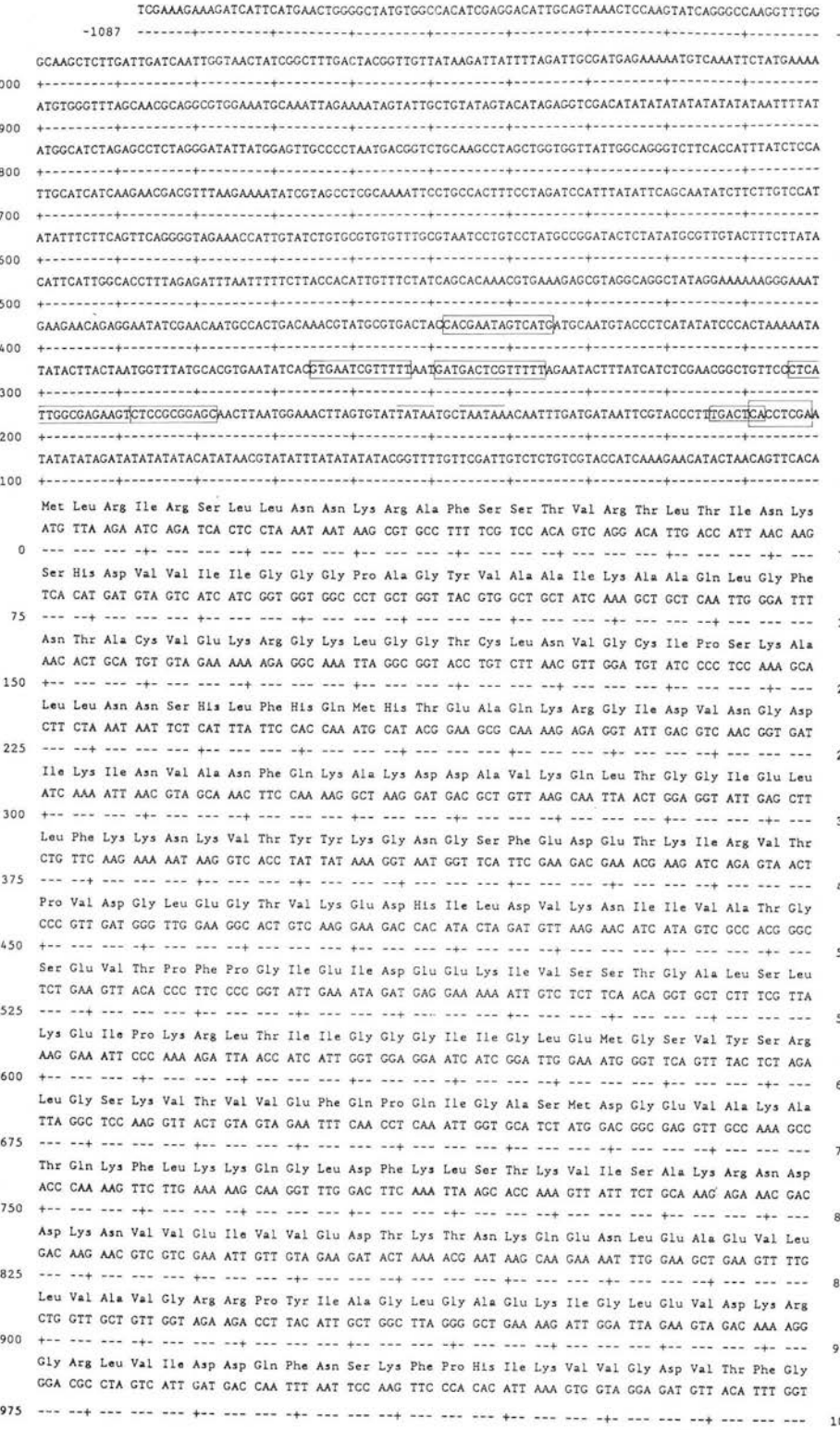
The primary structure, translated from the nucleotide sequence of the *LPD1* gene, contains 499 amino acid residues that correspond to a protein of M_r 54010 (54730 including the FAD cofactor). These M_r values are in good agreement with previous estimates (Wieland, 1983).

Lipoamide dehydrogenase, encoded in the nucleus and synthesized on cytoplasmic ribosomes, is a component of two multienzyme complexes which function in the mitochondrial matrix–inner membrane compartment. The presence of an N-terminal targeting sequence, which directs the protein to its correct subcellular location and facilitates its entry into the mitochondrion, has been shown in mammalian cells. Antibodies raised against the E1, E2 and E3 components of mitochondrial 2-oxoglutarate dehydrogenase from ox heart have been used to show the presence, in cultured pig kidney cells, of initial cytoplasmic translation products that are larger than the mature proteins (Hunter & Lindsay, 1986). The E3 component is synthesized as a polypeptide 10–20 amino acids larger than the mature protein. The primary structure, derived from the *LPD1* nucleotide sequence of *S. cerevisiae*, supports the conclusion that, in this organism, the lipoamide dehydrogenase N-terminal region represents a mitochondrial targeting sequence. The first 20 amino acids are rich in seryl, threonyl and basic residues, and show a complete absence of acidic residues. All of these features are typical of mitochondrial targeting sequences in yeast (von Heijne, 1986). In *E. coli* such a targeting sequence would be unnecessary; and when the primary structures of *E. coli* and *S. cerevisiae* lipoamide dehydrogenase (Stephens *et al.*, 1983) are aligned (see Fig. 2) the yeast sequence is seen to contain an additional 20 amino acids at its N-terminus which are absent from the inferred *E. coli* protein.

The primary structures of several flavoproteins have been determined either partially or completely. Pig heart lipoamide dehydrogenase, *E. coli* lipoamide dehydrogenase, human erythrocyte glutathione reductase and *Pseudomonas aeruginosa* transposon mercuric reductase amino acid sequences have previously been compared (Williams *et al.*, 1984) and four regions were identified which show strong homology between the four proteins. Two of these (FAD-1 and FAD-2) are involved in the binding of FAD, and the others constitute a pyridine nucleotide binding domain and an 'interface' domain involved in the interaction between subunits of the enzyme. Fig. 2 shows the primary structure of these four proteins and of *S. cerevisiae* lipoamide dehydrogenase aligned for maximum homology using the computer programs GAP and PRETTY. The primary structure of *S. cerevisiae* lipoamide dehydrogenase is in full agreement with the established 'domain homology' found between the other proteins. Comparison of the three lipoamide dehydrogenase primary structures reveals that the strongest homology lies between those of pig heart and yeast. The higher level of homology is particularly noticeable in the dimer interface region concerned with the interaction of the two polypeptides in formation of the active dimeric form (see Fig. 2). This may indicate that there are differences between prokaryotes and eukaryotes in the recognition of the subunits during assembly, which is reflected in the different organization of the multienzyme complexes in terms of subunit composition, especially for pyruvate dehydrogenase (Reed, 1974). Despite these differences the overall degree of homology that exists between all three polypeptides indicates the heavy evolutionary constraints placed upon an enzyme which functions within two different multienzyme complexes. This is also reflected in the low immunogenicity of the yeast E3 subunit relative to the other enzymes of the pyruvate and 2-oxoglutarate dehydrogenase complexes (De Marcucci *et al.*, 1985).

General features of the noncoding regions

The sequences upstream and downstream which flank the coding region of the *LPD1* gene share several characteristics common to the 5′- and 3′-noncoding regions of many sequenced yeast genes.



(*continued on facing page*)

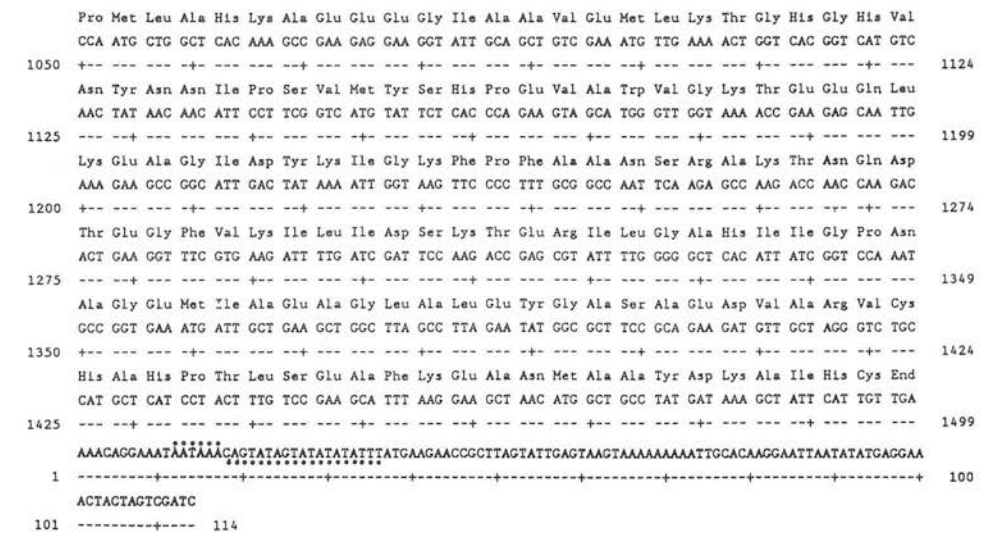


Fig. 1. Nucleotide sequence of the *LPD1* gene and deduced primary structure of lipoamide dehydrogenase. The nucleotide sequence of a 2·7 kb region containing the sense strand of the *LPD1* gene as well as over 1 kb of flanking DNA is shown in the 5′ to 3′ direction. The specific features of the 5′ flanking region are noted as follows: The 'TATA' consensus sequences are overlined; the sequences which show homology to known regulatory sequences are boxed. Specific features of the 3′ flanking sequence are noted as follows: the proposed consensus polyadenylation signal, AATAAA, is denoted by overdots; the proposed polyadenylation/termination signal TAG...TATGT...TTT is denoted by underdots.

The noncoding regions of several yeast genes are AT rich and the regions 5′ (−500 to −1) and 3′ to the *LPD1* coding region have overall A + T compositions of 63% and 74% respectively. The sequences TAATAA and TATAA found at positions −146 and −154 represent potential TATA boxes believed to be important in positioning transcription initiation by RNA polymerase II (Grosschedl & Birnstiel, 1980). An A is found at position −3 as has been reported for the majority of yeast genes sequenced (Kozak, 1981).

Downstream of the TGA translation termination codon an additional 114 bp has been sequenced. The motif AATAAA postulated to specify a site involved in polyadenylation in higher eukaryotes (Proudfoot & Brownlee, 1976) and seen in many yeast genes is found at position +12. The sequence TAG...TA(T)GT...(AT rich)...TTT has been postulated to be a signal for transcription termination or polyadenylation (Zaret & Sherman, 1982), and a version of this is found at position +18: CAGTATAGTATATATATTT.

Regulation of the LPD1 gene

The upstream region from −358 to −102 contains a number of motifs which show homology to ones that have roles in the transcriptional regulation of other yeast genes. This region therefore probably contains elements which control the expression of the *LPD1* gene.

These motifs are illustrated in Fig. 3. At position −247 there is a TGACTC sequence with an adjacent run of T residues that conforms very closely to the consensus sequence obtained from the GCN4 protein binding sites studied to date (Hill *et al.*, 1986). This protein mediates the general control of amino acid biosynthesis response that is involved in modulating during amino acid starvation the expression of some genes for amino acid synthesis (Struhl, 1982). It is interesting that a second, almost perfect direct repeat sequence (TGAATCGTTTTT), is also present 17 bp away at position −265. At −114 there is another TGACTC motif, although this conforms less well to the GCN4 consensus. All other yeast genes found so far to be subject to this GCN4-mediated control are concerned directly with synthesis of amino acids or charged tRNA species. Regulation of lipoamide dehydrogenase by general amino acid control may be less

```
          1                                                                  70
YLPDH     mlrirsllnn krafsstvrt ltinkshDVV IIGGGPaGYv AAikAAqLGF nTacVEkrgk LGGTCLNVGC
PLPDH     .......... .......... adqpidaDVt VIGsGPGGYv AAikAAqLGF kTVcIEkneT LGGTCLNVGC
ECLPDH    .......... .......... .steiktqVV VlGaGPaGYs AAfrcAdLGL eTViVErynT LGGvCLNVGC
HGR       ......acrq epqpqgpdpa agavasyDyl VIGGGsGGla sArrAAeLGa raavVE.shk LGGTCvNVGC
PAMR      adnrvglldk vrgwmaaaek hsgneppvqV aViGsgGaam AAalkAveqg aqVtliergT iGGTCvNVGC
          ---------- ---------- -------DVV VIGGGPGGY- AA--AA-LGF -TV-VE---T LGGTCLNVGC

                                                                             139
          IPSKALLNns HlfHqMhtea qkrgi.dvng DIKiNvanfq kakDdaVkqL TGGielLFKk nkVtyykGng
          IPSKALLNns HyyH.Mahgk dfasrGiems EVrlNlekm* ********aL TGGiahLFK* **********
          IPSKALL... HvakvieEak alaehGivfg EpKtdidkir twkEkvInqL TGGlagMaKg rkVkVVnGlg
          VPkKvMwNta vhsefMhDha ...dyGfpsc EgKfNwrvik ekrDayVsrL naiyqnnLtk shIeIIrGha
          VPSKiMiraa HiaHlrrEsp fdggiaatvp tIdrskllaq qqarvdelrh akyegiLggn paItVVhGea
          IPSKALLN-- H--H-M-E-- -----G---- EIK-N----- ---D--V--L TGG---LFK- ----VV-G--

                                                                             208
          sFedetkirV tpvdGlegtV kedhIldvkN IiVATG.Sev TPFPgIeiDE ekivsSTgAL sLkEIPkRLt
          ***vngygk* ********ad gsteVintkN IlIATG.Sev TPFPgItiDE ********** ********kMv
          kFtgantleV egenG..... ..ktVinfdN aiIAaG.Srp iqLPfIphED priwdSTdAL eLkEVPeRLl
          aF........ tsdpkptieV sgkkytaph. IlIATGgmps TPhesqipga slgitSdgfF qLeElPgRsv
          rFkddqsltV rlneGgervV mfdrcl.... ..VATG.asp avpPipglkE spywtSTeAL asdtIPeRLa
          -F-------V ----G----V ----V----N I-IATG-S-- TPFP-I--DE -----ST-AL -L-EIP-RL-

                                                                             278
          IIGgGIIGLE MgsVYsrLGS KVtVvefqpq IgasmDgeVa katqkflkkq gldFklstkv isakrnddkn
          VIGaGVIGvE ..******** ********** ********** ********** ********** **********
          VmGgGIIGLE MgtVYhaLGS qIdVvemfdq VipaaDkdIv kvftkriskk fnlMletkvt aveakedgiy
          IVGaGyIavE MagILsaLGS Ktslmirhdk VlrsfDsmIs tncteelena gvevlkfsqv kevkktlsgl
          VIGssVVaLE LaqaFarLGS KVtVlarntl ffredpaige avtaafraeg ievLehtqas qva.......
          VIG-G-IGLE M--VY--LGS KV-V------ V----D--I- ---------- ---------- ----------

                                                                             346
          vveiv..ved tktnkqenle aEvLLVAVGR rPyiagLgaE KiGlEVDkrG rlVIDdqFns kfPhIkvVGD
          ********** ********** ****LVcIGR ePfTqnLgLE elGIElr*** ********** *IPnIaAIGD
          v......tme gkkapaepqr yDavLVAIGR vPNgknLdag KaGVEVDdrG fIrVDkqLrT nVPhIFAIGD
          evsmvtavpg rlpvmtmipd vDcLLwAIGR vPNTkdLsLn KlGIqtDdkG hIIVDefqnT nVkgIYAVGD
          ....hmdgef vlttthgelr aDkLLVAtGR tPNTrsLaLD aaGVtVnaqG aIVIDqgMrT snPnIYAaGD
          ---------- ---------- -D-LLVAIGR -PNT--L-LE K-G-EVD--G -IV-D----T -VP-IYA-GD

                                                                             412
          VtfgPMLAHK AeeEG..iAa VemLktGhgh vnYNnIPSVm YsHPEVAWVG kT..EeqlKE aGIDYKigkF
          VvagPMLAHK AedEG..iic VegMagGavh iDYNcVPSVI YTHPEVAWVG k..sEeqlKE eGIEYKvgkF
          IvgqPMLAHK gvhEGhvaAe Viagk..khy FDpkvIPSIa YTePEVAWVG lT..EkeaKE kGIsYetatF
          VcgkaLLtpv AiaaGrklAh rlfeykedsk LDYNnIPtVV FsHPpIgtVG lTedEaihKy gienvKtyst
          ctdqPqFvyv AaaaG..tra ainMtgGdaa LDltamPaVV FTdPqVAtVG ..ysEaeahh dGIEtdsrtL
          V---PMLAHK A--EG---A- V--M--G--- LDYN-IPSVV YTHPEVAWVG -T--E---KE -GIEYK---F

                                                                             482
          PFAansRAkT NqDTEGfVKi liDsktERIl GaHIIGPnAG EMIaEAgLAl EyGAsaEDVA rvcHAHPTLS
          PFA***RAkT NaDTDGmVKi lgqkstDRVl GaHIIGPgAG EMInEAALAl EyGAscEDIA rvcHAHPTLS
          PWAasgRAia sdcaDGmtKL IfDkeshRVI GgaIVGtngG ELlgEigLAI EmGcdaEDIA ltiHAHPTLh
          sFtpmyhAvT krkTkcvmKM VcankeEkVV GiHmqGlgcd EMlqgfAvAV kmGAtkaDfd ntvaiHPTsS
          tLdnvpRAla NfDTrGfIKL ViEegshRlI GvqaVaPeAG ELIqtAALAI rnrmtvqElA dqlfpylTMv
          PFA---RA-T N-DTDG-VKL V-D---ERVI G-HI-GP-AG EMI-EAALAI E-GA--EDIA ---HAHPTLS

                    499
          EaFkeAnmaa ydkaihc... .....
          EaFreAnlaa sfgkainf.. .....
          EsvglAaevf egsitdlpnp kakkk
          EeLvtlr... .......... .....
          EgLklAaqtf nkdvkqlscc ag...
          E----A---- ---------- -----
```

(*a*)

GCN4 binding site consensus sequence		rrTGACTCattt---t	
LPD1	-264	CGTGAATCGTTTTTAG	-253
LPD1	-247	GATGACTCGTTTTTAA	-236
LPD1	-114	TTTGACTCACCTCGAA	-108

(*b*)

CYC1 UAS1		CTCTTTGGCCGGGGTTT	
CYC1 UAS2		CTCTTTGGCGAGCGTTT	
GAL1–10 control region sequence		CTCTTTGGAACTTTCAG	
LPD1	-204	CTCATTGGCGAGAAGTC	-189

(*c*)

Heat-shock consensus sequence		C--GAA--TTC--G	
LPD1	-348	CACGAATAGTCATG	-335

Fig. 3. Alignment of potential regulatory sequences with known motifs. In each case the established regulatory sequence is displayed, with the upstream regions of the *LPD1* gene which show homology shown directly below. (*a*) Consensus sequence for the general control regulatory site. Highly conserved nucleotides are shown as capitals, conserved residues are shown in lower case, and nonconserved residues are indicated by dashes. (*b*) Upstream activation sites (UAS1 and UAS2) of *CYC1* and a sequence located in the *GAL1–10* control region. (*c*) Heat-shock promoter consensus sequence.

surprising than it seems at first sight, since some of the tricarboxylic acid (TCA) cycle intermediates are substrates for the synthesis of such amino acids as glutamate, arginine, lysine and proline. There is one report that fumarase activity is derepressed in arginine bradytrophs; it has been suggested that this TCA cycle enzyme may also be under general amino acid control (Delforge *et al.*, 1975).

A second putative control region, at position −204, together with its homology with the upstream activation sites of *CYC1*, is shown in Fig. 3. The function of this region is not known but both genes are involved in aerobic metabolism and are subject to catabolite repression (Guarente *et al.*, 1984; Roy & Dawes, 1987); a related sequence with homology over 8 bases can be found in the *GAL1–10* control region (Johnston & Davis, 1984).

A third potential control site shown in Fig. 3 is the sequence at −352 which shows a 7 out of 8 match with the general heat-shock consensus sequence (Pelham, 1985).

Several further sequences are also of interest. One is an inverted repeat CTCCGCGGAG at position −188 which is similar to a general repressor protein binding sequence proposed by

Fig. 2. Protein sequence comparison. The primary structures of *S. cerevisiae* lipoamide dehydrogenase (YLPDH), pig heart lipoamide dehydrogenase (PLPDH), *E. coli* lipoamide dehydrogenase (ECLPDH), human red blood cell glutathione reductase (HGR) and *Pseudomonas aeruginosa* mercuric reductase (PAMR) are aligned for maximum homology. The residue numbering is based on *S. cerevisiae* lipoamide dehydrogenase. A consensus line below the five sequences shows residues common to three or more of the amino acid sequences. The borders of the domains representing the four regions of strong homology are based on glutathione reductase and are marked by arrowheads below the consensus sequence. The 80 residues at the N-terminus of mercuric reductase have been omitted. The individual sequences show residues in agreement with the consensus sequence in upper case and all others in lower case. The alignment was performed using the program PRETTY with the default protein comparison file (Gribskov & Burgess, 1986), which regards certain amino acids as similar for the purposes of the consensus line. The asterisks in the pig heart lipoamide dehydrogenase sequence indicate regions of unknown sequence. Residues common to pig heart and yeast lipoamide dehydrogenase but not to the *E. coli* lipoamide dehydrogenase are shown by circles above the yeast sequence.

T. Cooper (personal communication); a second is the sequence CACCTCGA at position −109 which is homologous to one found upstream of the *ARG3* and *CAR1* genes involved in arginine biosynthesis and degradation respectively (Crabeel *et al.*, 1985). In addition to the above, the upstream region contains a number of direct repeats and inverted repeats which may also have a role in the regulation of *LPD1* expression. One of these at −283 is very interesting since it is a 23 bp inverted repeat, containing direct repeats of 9 bp each within it. It is represented in a symmetrically truncated form at −361. We are currently investigating the functional significance of these upstream regions.

We are grateful to Lynn Kennedy for technical assistance, and to the SERC for a postgraduate studentship (J.R.) and research grant support.

REFERENCES

CRABEEL, M., HUYGEN, R., VERSCHUEREN, K., MESSENGUY, F., TINEL, K., CUNIN, R. & GLANSDORFF, N. (1985). General amino acid control and specific arginine repression in *Saccharomyces cerevisiae*: physical study of the bifunctional regulatory region of the *ARG3* gene. *Molecular and Cellular Biology* **5**, 3139–3148.

DELFORGE, J., MESSENGUY, F. & WIAME, J. (1975). The regulation of arginine biosynthesis in *Saccharomyces cerevisiae*. *European Journal of Biochemistry* **57**, 231–239.

DE MARCUCCI, O. L., HUNTER, A. & LINDSAY, J. G. (1985). Low immunogenicity of the common lipoamide dehydrogenase subunit (E3) of mammalian pyruvate dehydrogenase and 2-oxoglutarate dehydrogenase multienzyme complexes. *Biochemical Journal* **226**, 509–517.

DEVEREUX, J., HAEBERLI, P. & SMITHIES, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Research* **12**, 387–395.

DICKINSON, J. R., ROY, D. J. & DAWES, I. W. (1986). A mutation affecting lipoamide dehydrogenase, pyruvate dehydrogenase and 2-oxoglutarate dehydrogenase activities in *Saccharomyces cerevisiae*. *Molecular and General Genetics* **204**, 103–107.

GRIBSKOV, M. & BURGESS, R. R. (1986). Sigma factors from *E. coli*, *B. subtilis*, phage SP01 and phage T4 are homologous proteins. *Nucleic Acids Research* **14**, 6745–6763.

GROSSCHEDL, R. & BIRNSTIEL, M. L. (1980). Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants *in vivo*. *Proceedings of the National Academy of Sciences of the United States of America* **77**, 1432–1436.

GUARENTE, L., LALONDE, B., GIFFORD, P. & ALANI, E. (1984). Distinctly regulated tandem upstream activation sites mediate catabolite repression of the *CYC1* gene of *Saccharomyces cerevisiae*. *Cell* **36**, 503–511.

GUEST, J. R. & CREAGHAN, I. T. (1973). Gene–protein relationships of the α-keto acid dehydrogenase complexes of *Escherichia coli* K12: isolation and characterization of lipoamide dehydrogenase mutants. *Journal of General Microbiology* **75**, 197–210.

GUEST, J. R. & RICE, D. W. (1984). Molecular genetic approaches to the study of *E. coli* flavoproteins. In *Flavins and Flavoproteins*, pp. 111–123. Edited by R. C. Bray. Berlin: Walter de Gruyter.

VON HEIJNE, G. (1986). Mitochondrial targeting sequences may form amphiphilic helices. *EMBO Journal* **5**, 1335–1342.

HILL, D. E., HOPE, I. A., MACKE, J. P. & STRUHL, K. (1986). Saturation mutagenesis of the yeast *his3* regulatory site: requirements for transcriptional induction and for binding by GCN4 activator protein. *Science* **234**, 451–457.

HUNTER, A. & LINDSAY, J. G. (1986). Immunological and biosynthetic studies on the mammalian 2-oxoglutarate dehydrogenase multienzyme complex. *European Journal of Biochemistry* **155**, 103–109.

JOHNSTON, M. & DAVIS, R. W. (1984). Sequences that regulate the divergent *GAL1–GAL10* promoter in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology* **4**, 1440–1448.

KOCHI, H. & KIKUCHI, G. (1976). Mechanism of reversible glycine cleavage reaction in *Arthrobacter globiformis*: function of lipoic acid in the cleavage and synthesis of glycine. *Archives of Biochemistry and Biophysics* **173**, 71–81.

KOZAK, M. (1981). Possible role of flanking nucleotides in recognition of the AUG initiation codon by eukaryotic ribosomes. *Nucleic Acids Research* **9**, 5233–5252.

LAWSON, U. K., COOK, K. G. & YEAMAN, S. J. (1983). Rapid purification of bovine kidney branched-chain 2-oxoacid dehydrogenase complex containing endogenous kinase activity. *FEBS Letters* **157**, 54–58.

MANIATIS, T., FRITSCH, E. F. & SAMBROOK, J. (1982). *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

MESSING, J. (1983). New M13 vectors for cloning. *Methods in Enzymology* **101**, 20–78.

MUKHERJEE, B. B., MATTHEWS, J., HORNEY, O. L. & REED, L. J. (1965). Resolution and reconstitution of *Escherichia coli* α-ketoglutarate dehydrogenase complex. *Journal of Biological Chemistry* **240**, 2268–2269.

PELHAM, H. (1985). Activation of heat shock genes in eukaryotes. *Trends in Genetics* **1**, 31–35.

PROUDFOOT, N. J. & BROWNLEE, G. G. (1976). 3′ noncoding region sequences in eukaryotic messenger RNA. *Nature, London* **263**, 211–214.

REED, L. J. (1974). Multienzyme complexes. *Accounts in Chemical Research* **7**, 40–46.

ROBINSON, J. R., KLEIN, S. M. & SAGERS, R. D. (1973). Glycine metabolism. Lipoic acid as the prosthetic

group in the electron transfer protein P^2 from *Peptococcus glycinophilus*. *Journal of Biological Chemistry* **248**, 5319–5323.

ROY, D. J. & DAWES, I. W. (1987). Cloning and characterization of the gene encoding lipoamide dehydrogenase in *Saccharomyces cerevisiae*. *Journal of General Microbiology* **133**, 925–933.

SANGER, F., NICKLEN, S. & COULSEN, A. R. (1977). DNA sequencing with chain terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 5463–5467.

SMITH, H. O. (1980). Recovery of DNA from gels. *Methods in Enzymology* **65**, 371–380.

SPENCER, M. E. & GUEST, J. R. (1985). Transcriptional analysis of the *suc*AB *ace*EF and *lpd* genes of *E. coli*. *Molecular and General Genetics* **200**, 145–154.

STEPHENS, P. E., LEWIS, H. M., DARLISON, M. G. & GUEST, J. R. (1983). Nucleotide sequence of the lipoamide dehydrogenase gene of *Escherichia coli* K12. *European Journal of Biochemistry* **135**, 519–527.

STRUHL, K. (1982). Regulatory sites for *HIS3* expression in yeast. *Nature, London* **300**, 284.

WIELAND, O. H. (1983). The mammalian pyruvate dehydrogenase complex: structure and regulation. *Physiological and Biochemical Pharmacology* **96**, 123–170.

WILLIAMS, C. H. JR, ARSCOTT, L. D. & SWENSON, P. (1984). Active site chemical modification and sequencing of flavoproteins. In *Flavins and Flavoproteins*, pp. 95–109. Edited by R. C. Bray. Berlin: Walter de Gruyter.

WINTER, G. & FIELDS, S. (1980). Cloning of influenza cDNA into M13: the sequence of the RNA segment encoding the A/PR/8/34 matrix protein. *Nucleic Acids Research* **8**, 1965–1974.

ZARET, K. S. & SHERMAN, F. (1982). DNA sequences required for efficient transcription termination in yeast. *Cell* **28**, 563–573.